

TEXTBOOK OF GENERAL BACTERIOLOGY

TEXTBOOK OF GENERAL BACTERIOLOGY

DAVID WHITE



SPRINGER Med Press
www.springermedpress.com



SPRINGER Med Press

Published by Springer Med Press
1202 N. Orange Street Suit #600
Wilmington, DE19801 USA

© 2024 by Springer Med Press

ISBN: 979-8-88626-098-4

Textbook of General Bacteriology

David White

This work is subject to copyright. All rights are reserved by the publisher. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without prior permission in writing from the publisher.

Notice

Practitioners and researchers must always rely on their experience and knowledge in evaluating and using information, methods, compounds, or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by Springer Med Press, authors, editors, or contributors for any injury and damage to persons or property as a matter of products liability, negligence, or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from Library of Congress

For more information regarding Springer Med Press and its products, please visit the publisher's website
www.springermedpress.com

CONTENTS

Preface

xiii

1.	Microbial Metabolism	1
	Introduction	1
	Types	1
	Heterotrophic Microbial Metabolism	2
	Fermentation	2
	Special Metabolic Properties	3
	Methylotrophy	3
	Syntrophy	4
	Anaerobic Respiration	4
	Denitrification – nitrate as Electron Acceptor	4
	Sulfate Reduction – sulfate as Electron Acceptor	5
	Acetogenesis – carbon Dioxide as Electron Acceptor	5
	Other Inorganic Electron Acceptors	6
	Organic Terminal Electron Acceptors	6
	Chemolithotrophy	6
	Hydrogen Oxidation	6
	Sulfur Oxidation	7
	Ferrous Iron (Fe) Oxidation	7
	Nitrification	7
	Anammox	8
	Phototrophy	8
	Nitrogen Fixation	9
	Non-biological Natural Nitrogen Fixation	9
	Biological Nitrogen Fixation	9
	Industrial Nitrogen Fixation	11
2.	Genetics	15
	History	15
	Pre-mendelian Ideas on Heredity	16
	Mendel	17
	Post-mendel, Pre-rediscovery	18
	Rediscovery of Mendel	19
	Emergence of Molecular Genetics	19
	Early Timeline	20
	The DNA Era	22
	The Genomics Era	24
	Features of Inheritance	26
	Discrete Inheritance and Mendel's Laws	26
	Notation and Diagrams	26
	Multiple Gene Interactions	27
	Molecular Basis for Inheritance	27
	DNA and Chromosomes	27
	Reproduction	28
	Recombination and Genetic Linkage	28
	Gene Expression	29

Genetic Code	29
Nature and Nurture	29
Gene Regulation	30
Genetic Change	30
Mutations	30
Natural Selection and Evolution	41
Model Organisms	42
Medicine	42
Research Methods	43
DNA Sequencing and Genomics	43
Society and Culture	44

3. **Human Microbiota** 45

Terminology	45
Relative Numbers	45
Study	46
Types	46
Bacteria	46
Archaea	47
Fungi	47
Viruses	47
Anatomical Areas	47
Skin	47
Conjunctiva	53
Gut	53
Vagina	63
Placenta	65
Uterus	67
Oral Cavity	68
Lung	70
Biliary Tract	72
Disease and Death	72
Cancer	72
Inflammatory Bowel Disease	73
Human Immunodeficiency Virus	73
Death	73
Environmental Health	74

4. **Pathogenesis** 75

Introduction	75
Description	75
Acute (Medicine)	75
Meaning Variations	76
Acute Care	76
Chronic Condition	76
Types	76
Risk Factors	78
Prevention	78
Epidemiology	78
Economic Impact	79
Relapse	80
Causes	80
Risk Factors	81
Triggers	82
Treatment	82
Animal Models	91
Differences between Sexes	92
Causes	93
Types	93
Vascular Component	95
Cellular Component	97

Morphologic Patterns	99
Inflammatory Disorders	99
Systemic Effects	103
Outcomes	105
Diet and Inflammation	105
Exercise and Inflammation	106
Malignancy	108
Necrosis	109
Causes	110
Pathogenesis	111
Treatment	111
In Plants	112
Immune System	112
History of Immunology	113
Layered Defence	113
Innate Immune System	113
Adaptive Immune System	116
Disorders of Human Immunity	119
Other Mechanisms and Evolution	120
Tumor Immunology	120
Physiological Regulation	121
Manipulation in Medicine	123
Theoretical Approaches to the Immune System	123
Predicting Immunogenicity	123
Manipulation by Pathogens	124

5. Adaptive Immune System	125
Introduction	125
Functions	126
Lymphocytes	126
Development	127
Characteristics	127
Lymphocytes and Disease	127
T lymphocytes	128
CD8+ T Lymphocytes and Cytotoxicity	128
Helper T-cells	129
Gamma Delta T cells	129
Gene Families in Different Species	130
B lymphocytes and Antibody Production	132
Alternative Adaptive Immune System	133
Immunological Memory	133
Passive Memory	133
Active Memory	134
Immunological Diversity	134
Adaptive Immunity During Pregnancy	135
Immune Network Theory	135
The Symmetrical Immune Network Theory	136
I-J, the I-J Paradox, and a Resolution of the I-J Paradox	136
Relevance for Understanding HIV Pathogenesis	136
An HIV Vaccine Concept based on Immune Network Theory	137
A more General Vaccine based on Immune Network Theory	137
MHC Restriction of V-V Interactions in Serum IGG	138
Evolution	138
Types of Adaptive Immunity	138
Antigens	139
Terminology	140
Sources	140
Antigenic Specificity	142
Superantigen	142
Allergen	146
Hapten	150

Epitope	152
Mimotope	154
Antigen Presentation	155
Antibodies	171
Forms	172
Antibody–antigen Interactions	173
Isotypes	173
Structure	174
Function	175
Immunoglobulin Diversity	177
Medical Applications	179
Research Applications	180
Regulations	181
Structure Prediction	181
Immune Complex	182

6. **Epidemiology** **183**

History	183
Modern Era	183
Types of Studies	185
Case Series	185
Case-control Studies	185
Cohort Studies	186
Causal Inference	187
Bradford Hill Criteria	187
Legal Interpretation	188
Population-based Health Management	188
Applied Field Epidemiology	189
Humanitarian Context	189
Validity: Precision and Bias	189
Random Error	190
Systematic Error	190
The Profession	191
Transmission (Medicine)	192
Definition and Related Terms	192
Routes	192
Tracking	210
Relationship with Virulence and Survival	210
Beneficial Microorganisms	210
Cross-species Transmission	211
Epidemiological Transition	215

7. **Medical Diagnosis** **221**

Introduction	221
History	221
Medical Uses	221
Procedure	222
Adverse Effects	223
Society and Culture	224
History of Medical Diagnosis	226
Ancient Egypt	226
Ancient Babylonia	226
Ancient China	226
Ancient Greece	226
Islamic World	226
Middle Ages	227
19th Century	227
The Oslerian Ideal	227
Garrod's View	227
Present-day Oslerian Practice	228
Influence of DNA Technology	228

Photoacoustic Effect	228
Overdiagnosis	228
Nucleic Acid Test	229
Nekrep Case	229
Antisaccade Task	229
Procedure	230
History	230
Current Use	230
Body Fluid	231
List of Body Fluids	231
Health	232
Body Fluids in Art	232
Canards Sign	233
CIT Programme Tumor Identity Cards	233
Companion Diagnostic	233
Coordinated Diagnostics	234
Practice	234
Industry Applicability	234
Cosmed	234
History	234
Products	235
Appearances	235
Awards	235
Cryoglobulinemia	235
Classification	235
Signs and Symptoms	236
Pathophysiology	237
Diagnosis	238
Treatment	238
Diagnostic Overshadowing	240
Differential Diagnosis	240
General Components	240
Specific Methods	241
Coverage of Candidate Conditions	241
Combinations	242
Machine Differential Diagnosis	242
History	242
Alternative Medical Meanings	243
Usage Apart from in Medicine	243
Direct Hit	243
Technology	243
Clinical Trials	243
Cost	244
Dot Cancellation Test	244
Drehmann Sign	244
Elbow Extension Test	244
Administering the Test	244
Utility	244
Role of Electrophysiology in Clinical Medicine	245
Electrophysiological Techniques	245
Diagnosable Conditions and Applicable Methods	246
External Cause	248
FAST (Stroke)	248
History	249
Fine-needle Aspiration	249
Medication Uses	249
Preparation	249
Post-operative Care and Complications	250
Criticism	250
Flow-mediated Dilation	250
Clinical Significance	251

Effects of Exercise	251
Frax	251
Components	251
Availability and Usage	252
Gender-bias in Medical Diagnosis	252
Medical Diagnosis	252
Psychological Diagnosis	252
Immunolabeling	254
Indirect vs. Direct Method	254
Antibody Binding and Specificity	255
Specific Immunolabeling Techniques	255
Instantaneous Wave-free Ratio	257
Assessment of Coronary Stenosis	257
IFR Definition and Calculation	257
Studies	258
IFR Clinical Outcome Studies	258
IFR Physiological Mapping of Coronary Vessels	259
Controversy	259
Laboratory Diagnosis of Viral Infections	259
Sampling	259
Virus Isolation	260
Nucleic Acid based Methods	260
Microscopy based Methods	261
Host Antibody Detection	261
Hemagglutination Assay	261
Martha Mitchell Effect	261
Description	261
Origin	262
Medical Error	262
Definitons	262
Impact	262
Causes	263
Examples	265
After an Error has Occurred	265
Prevention	268
Misconceptions	269
Medical Findings	269
Definition	269
Types of Findings	270
Findings and Symptoms	270
Medical Laboratory	270
Departments	270
Medical Laboratory Staff	271
Types of Laboratory	272
Specimen Processing and Work Flow	272
Laboratory Informatics	273
Result Analysis, Validation and Interpretation	273
Medical Laboratory Accreditation	273
Mental Status Examination	274
Theoretical Foundations	274
Application	274
Domains	275
Cultural Considerations	279
Children	280
Microvesicles	280
Formation and Contents	280
Role on Target Cells	282
Relevance in Disease	283
Clinical Applications	285
Sutton's Law	287
Surgical Sieve	287

Examples	288
Pathologic basis of Disease	288
<i>Bibliography</i>	289
<i>Index</i>	293

PREFACE

This book aims at giving a comprehensive survey of our present knowledge of bacteriology, especially in its relation to medical work. The book gives promise of a very valuable contribution to scientific literature. Author can have only a superficial knowledge of many of the articles, and his judgment of the whole work must, to a certain extent, be influenced by the quality of those papers with which he has an intimate knowledge. It deals with all aspects of plague infection-the characters of the bacillus, the problems in plague immunity, the transmission by the rat-flea, the natural disease in rodents, and the various forms of plague in man with reference to its pathology, its diagnosis, and treatment. It is enrich with valuable contributions and regarded as a standard work on the subject.

The author would like to express his sincere thanks to his friends and family for their support in the completion of this book. He wants to thank Springer MedPress Corporation to recognise the intensity of the subject and help me publish this masterpiece which will educate well the newcomers. Also thanks the staff and the editorial team for their kind support and encouragement. This book has all the relevant details with the new concepts.

-Author

Microbial Metabolism

INTRODUCTION

Microbial metabolism is the means by which a microbe obtains the energy and nutrients (*e.g.* carbon) it needs to live and reproduce. Microbes use many different types of metabolic strategies and species can often be differentiated from each other based on metabolic characteristics. The specific metabolic properties of a microbe are the major factors in determining that microbe's ecological niche, and often allow for that microbe to be useful in industrial processes or responsible for biogeochemical cycles.

TYPES

All microbial metabolisms can be arranged according to three principles:

1. How the organism obtains carbon for synthesising cell mass:
 - Autotrophic – carbon is obtained from carbon dioxide (CO_2)
 - Heterotrophic – carbon is obtained from organic compounds
 - Mixotrophic – carbon is obtained from both organic compounds and by fixing carbon dioxide
2. How the organism obtains reducing equivalents used either in energy conservation or in biosynthetic reactions:
 - Lithotrophic – reducing equivalents are obtained from inorganic compounds
 - Organotrophic – reducing equivalents are obtained from organic compounds
3. How the organism obtains energy for living and growing:
 - Chemotrophic – energy is obtained from external chemical compounds
 - Phototrophic – energy is obtained from light

In practice, these terms are almost freely combined. Typical examples are as follows:

- Chemolithoautotrophs obtain energy from the oxidation of inorganic compounds and carbon from the fixation of carbon dioxide. Examples: Nitrifying bacteria, Sulfur-oxidizing bacteria, Iron-oxidizing bacteria, Knallgas-bacteria
- Photolithoautotrophs obtain energy from light and carbon from the fixation of carbon dioxide, using reducing equivalents from inorganic compounds. Examples: Cyanobacteria (water (H_2O) as reducing equivalent donor), Chlorobiaceae, Chromatiaceae (hydrogen sulfide (H_2S) as reducing equivalent donor), *Chloroflexus* (hydrogen (H_2) as reducing equivalent donor)
- Chemolithoheterotrophs obtain energy from the oxidation of inorganic compounds, but cannot fix carbon dioxide (CO_2). Examples: some *Thiobacillus*, some *Beggiatoa*, some *Nitrobacter* spp., *Wolinella* (with H_2 as reducing equivalent donor), some Knallgas-bacteria, some sulfate-reducing bacteria
- Chemoorganoheterotrophs obtain energy, carbon, and reducing equivalents for biosynthetic reactions from organic compounds. Examples: most bacteria, *e.g.* *Escherichia coli*, *Bacillus* spp., *Actinobacteria*

- Photoorganoheterotrophs obtain energy from light, carbon and reducing equivalents for biosynthetic reactions from organic compounds. Some species are strictly heterotrophic, many others can also fix carbon dioxide and are mixotrophic. Examples: *Rhodobacter*, *Rhodopseudomonas*, *Rhodospirillum*, *Rhodomicrobium*, *Rhodocyclus*, *Heliobacterium*, *Chloroflexus* (alternatively to photolithoautotrophy with hydrogen)

HETEROTROPHIC MICROBIAL METABOLISM

Some microbes are heterotrophic (more precisely chemoorganoheterotrophic), using organic compounds as both carbon and energy sources. Heterotrophic microbes live off of nutrients that they scavenge from living hosts (as commensals or parasites) or find in dead organic matter of all kind (saprophages). Microbial metabolism is the main contribution for the bodily decay of all organisms after death. Many eukaryotic microorganisms are heterotrophic by predation or parasitism, properties also found in some bacteria such as *Bdellovibrio* (an intracellular parasite of other bacteria, causing death of its victims) and Myxobacteria such as *Myxococcus* (predators of other bacteria which are killed and lysed by cooperating swarms of many single cells of Myxobacteria). Most pathogenic bacteria can be viewed as heterotrophic parasites of humans or the other eukaryotic species they affect. Heterotrophic microbes are extremely abundant in nature and are responsible for the breakdown of large organic polymers such as cellulose, chitin or lignin which are generally indigestible to larger animals. Generally, the breakdown of large polymers to carbon dioxide (mineralization) requires several different organisms, with one breaking down the polymer into its constituent monomers, one able to use the monomers and excreting simpler waste compounds as by-products, and one able to use the excreted wastes. There are many variations on this theme, as different organisms are able to degrade different polymers and secrete different waste products. Some organisms are even able to degrade more recalcitrant compounds such as petroleum compounds or pesticides, making them useful in bioremediation.

Biochemically, prokaryotic heterotrophic metabolism is much more versatile than that of eukaryotic organisms, although many prokaryotes share the most basic metabolic models with eukaryotes, e. g. using glycolysis (also called EMP pathway) for sugar metabolism and the citric acid cycle to degrade acetate, producing energy in the form of ATP and reducing power in the form of NADH or quinols. These basic pathways are well conserved because they are also involved in biosynthesis of many conserved building blocks needed for cell growth (sometimes in reverse direction). However, many bacteria and archaea utilize alternative metabolic pathways other than glycolysis and the citric acid cycle. A well-studied example is sugar metabolism via the keto-deoxy-phosphogluconate pathway (also called ED pathway) in *Pseudomonas*. Moreover, there is a third alternative sugar-catabolic pathway used by some bacteria, the pentose phosphate pathway. The metabolic diversity and ability of prokaryotes to use a large variety of organic compounds arises from the much deeper evolutionary history and diversity of prokaryotes, as compared to eukaryotes. It is also noteworthy that the mitochondrion, the small membrane-bound intracellular organelle that is the site of eukaryotic energy metabolism, arose from the endosymbiosis of a bacterium related to obligate intracellular *Rickettsia*, and also to plant-associated *Rhizobium* or *Agrobacterium*. Therefore, it is not surprising that all mitochondriate eukaryotes share metabolic properties with these Proteobacteria. Most microbes respire (use an electron transport chain), although oxygen is not the only terminal electron acceptor that may be used. As discussed below, the use of terminal electron acceptors other than oxygen has important biogeochemical consequences.

FERMENTATION

Fermentation is a specific type of heterotrophic metabolism that uses organic carbon instead of oxygen as a terminal electron acceptor. This means that these organisms do not use an electron transport chain to oxidize NADH to NAD and therefore must have an alternative method of using this reducing power and maintaining a

supply of NAD for the proper functioning of normal metabolic pathways (e.g. glycolysis). As oxygen is not required, fermentative organisms are anaerobic. Many organisms can use fermentation under anaerobic conditions and aerobic respiration when oxygen is present. These organisms are facultative anaerobes. To avoid the overproduction of NADH, obligately fermentative organisms usually do not have a complete citric acid cycle. Instead of using an ATP synthase as in respiration, ATP in fermentative organisms is produced by substrate-level phosphorylation where a phosphate group is transferred from a high-energy organic compound to ADP to form ATP. As a result of the need to produce high energy phosphate-containing organic compounds (generally in the form of Coenzyme A-esters) fermentative organisms use NADH and other cofactors to produce many different reduced metabolic by-products, often including hydrogen gas (H_2). These reduced organic compounds are generally small organic acids and alcohols derived from pyruvate, the end product of glycolysis. Examples include ethanol, acetate, lactate, and butyrate. Fermentative organisms are very important industrially and are used to make many different types of food products. The different metabolic end products produced by each specific bacterial species are responsible for the different tastes and properties of each food. Not all fermentative organisms use substrate-level phosphorylation. Instead, some organisms are able to couple the oxidation of low-energy organic compounds directly to the formation of a proton (or sodium) motive force and therefore ATP synthesis. Examples of these unusual forms of fermentation include succinate fermentation by *Propionigenium modestum* and oxalate fermentation by *Oxalobacter formigenes*. These reactions are extremely low-energy yielding. Humans and other higher animals also use fermentation to produce lactate from excess NADH, although this is not the major form of metabolism as it is in fermentative microorganisms.

SPECIAL METABOLIC PROPERTIES

Methylotrophy

Methylotrophy refers to the ability of an organism to use C1-compounds as energy sources. These compounds include methanol, methyl amines, formaldehyde, and formate. Several other less common substrates may also be used for metabolism, all of which lack carbon-carbon bonds. Examples of methylotrophs include the bacteria *Methylomonas* and *Methylobacter*. Methanotrophs are a specific type of methylotroph that are also able to use methane (CH_4) as a carbon source by oxidizing it sequentially to methanol (CH_3OH), formaldehyde (CH_2O), formate ($HCOO$), and carbon dioxide CO_2 initially using the enzyme methane monooxygenase. As oxygen is required for this process, all (conventional) methanotrophs are obligate aerobes. Reducing power in the form of quinones and NADH is produced during these oxidations to produce a proton motive force and therefore ATP generation. Methylotrophs and methanotrophs are not considered as autotrophic, because they are able to incorporate some of the oxidized methane (or other metabolites) into cellular carbon before it is completely oxidized to CO_2 (at the level of formaldehyde), using either the serine pathway (*Methylosinus*, *Methylocystis*) or the ribulose monophosphate pathway (*Methylococcus*), depending on the species of methylotroph. In addition to aerobic methylotrophy, methane can also be oxidized anaerobically. This occurs by a consortium of sulfate-reducing bacteria and relatives of methanogenic Archaea working syntrophically. Little is currently known about the biochemistry and ecology of this process. Methanogenesis is the biological production of methane. It is carried out by methanogens, strictly anaerobic Archaea such as *Methanococcus*, *Methanocaldococcus*, *Methanobacterium*, *Methanothermus*, *Methanosarcina*, *Methanosaeta* and *Methanopyrus*. The biochemistry of methanogenesis is unique in nature in its use of a number of unusual cofactors to sequentially reduce methanogenic substrates to methane, such as coenzyme M and methanofuran.

These cofactors are responsible (among other things) for the establishment of a proton gradient across the outer membrane thereby driving ATP synthesis. Several types of methanogenesis occur, differing in the starting compounds oxidized. Some methanogens reduce carbon dioxide (CO_2) to methane (CH_4) using electrons (most

often) from hydrogen gas (H_2) chemolithoautotrophically. These methanogens can often be found in environments containing fermentative organisms. The tight association of methanogens and fermentative bacteria can be considered to be syntrophic because the methanogens, which rely on the fermentors for hydrogen, relieve feedback inhibition of the fermentors by the build-up of excess hydrogen that would otherwise inhibit their growth. This type of syntrophic relationship is specifically known as interspecies hydrogen transfer. A second group of methanogens use methanol (CH_3OH) as a substrate for methanogenesis. These are chemoorganotrophic, but still autotrophic in using CO_2 as only carbon source. The biochemistry of this process is quite different from that of the carbon dioxide-reducing methanogens. Lastly, a third group of methanogens produce both methane and carbon dioxide from acetate (CH_3COO) with the acetate being split between the two carbons. These acetate-cleaving organisms are the only chemoorganoheterotrophic methanogens. All autotrophic methanogens use a variation of the reductive acetyl-CoA pathway to fix CO_2 and obtain cellular carbon.

SYNTROPHY

Syntrophy, in the context of microbial metabolism, refers to the pairing of multiple species to achieve a chemical reaction that, on its own, would be energetically unfavorable. The best studied example of this process is the oxidation of fermentative end products (such as acetate, ethanol and butyrate) by organisms such as *Syntrophomonas*. Alone, the oxidation of butyrate to acetate and hydrogen gas is energetically unfavorable. However, when a hydrogenotrophic (hydrogen-using) methanogen is present the use of the hydrogen gas will significantly lower the concentration of hydrogen (down to 10 atm) and thereby shift the equilibrium of the butyrate oxidation reaction under standard conditions (ΔG°) to non-standard conditions ($\Delta G'$). Because the concentration of one product is lowered, the reaction is “pulled” towards the products and shifted towards net energetically favorable conditions (for butyrate oxidation: $\Delta G^\circ = +48.2$ kJ/mol, but $\Delta G' = -8.9$ kJ/mol at 10 atm hydrogen and even lower if also the initially produced acetate is further metabolized by methanogens). Conversely, the available free energy from methanogenesis is lowered from $\Delta G^\circ = -131$ kJ/mol under standard conditions to $\Delta G' = -17$ kJ/mol at 10 atm hydrogen. This is an example of intraspecies hydrogen transfer. In this way, low energy-yielding carbon sources can be used by a consortium of organisms to achieve further degradation and eventual mineralization of these compounds. These reactions help prevent the excess sequestration of carbon over geologic time scales, releasing it back to the biosphere in usable forms such as methane and CO_2 .

ANAEROBIC RESPIRATION

While aerobic organisms during respiration use oxygen as a terminal electron acceptor, anaerobic organisms use other electron acceptors. These inorganic compounds have a lower reduction potential than oxygen, meaning that respiration is less efficient in these organisms and leads to slower growth rates than aerobes. Many facultative anaerobes can use either oxygen or alternative terminal electron acceptors for respiration depending on the environmental conditions. Most respiring anaerobes are heterotrophs, although some do live autotrophically. All of the processes described below are dissimilative, meaning that they are used during energy production and not to provide nutrients for the cell (assimilative). Assimilative pathways for many forms of anaerobic respiration are also known.

DENITRIFICATION – NITRATE AS ELECTRON ACCEPTOR

Denitrification is the utilization of nitrate (NO_3) as a terminal electron acceptor. It is a widespread process that is used by many members of the Proteobacteria. Many facultative anaerobes use denitrification because nitrate, like oxygen, has a high reduction potential. Many denitrifying bacteria can also use ferric iron (Fe) and some organic electron acceptors. Denitrification involves the stepwise reduction of nitrate to nitrite (NO_2),

nitric oxide (NO), nitrous oxide (N₂O), and dinitrogen (N₂) by the enzymes nitrate reductase, nitrite reductase, nitric oxide reductase, and nitrous oxide reductase, respectively. Protons are transported across the membrane by the initial NADH reductase, quinones, and nitrous oxide reductase to produce the electrochemical gradient critical for respiration. Some organisms (*e.g. E. coli*) only produce nitrate reductase and therefore can accomplish only the first reduction leading to the accumulation of nitrite. Others (*e.g. Paracoccus denitrificans* or *Pseudomonas stutzeri*) reduce nitrate completely. Complete denitrification is an environmentally significant process because some intermediates of denitrification (nitric oxide and nitrous oxide) are important greenhouse gases that react with sunlight and ozone to produce nitric acid, a component of acid rain. Denitrification is also important in biological wastewater treatment where it is used to reduce the amount of nitrogen released into the environment thereby reducing eutrophication.

SULFATE REDUCTION – SULFATE AS ELECTRON ACCEPTOR

Dissimilatory sulfate reduction is a relatively energetically poor process used by many Gram-negative bacteria found within the deltaproteobacteria, Gram-positive organisms relating to *Desulfotomaculum* or the archaeon *Archaeoglobus*. Hydrogen sulfide (H₂S) is produced as a metabolic end product. For sulfate reduction electron donors and energy are needed.

Electron Donors

Many sulfate reducers are organotrophic, using carbon compounds such as lactate and pyruvate (among many others) as electron donors, while others are lithotrophic, using hydrogen gas (H₂) as an electron donor. Some unusual autotrophic sulfate-reducing bacteria (*e.g. Desulfotignum phosphitoxidans*) can use phosphite (HPO₃) as an electron donor whereas others (*e.g. Desulfovibrio sulfodismutans*, *Desulfocapsa thiozymogenes*, *Desulfocapsa sulfoexigens*) are capable of sulfur disproportionation (splitting one compound into two different compounds, in this case an electron donor and an electron acceptor) using elemental sulfur (S), sulfite (SO₃), and thiosulfate (S₂O₃) to produce both hydrogen sulfide (H₂S) and sulfate (SO₄).

Energy for Reduction

All sulfate-reducing organisms are strict anaerobes. Because sulfate is energetically stable, before it can be metabolized it must first be activated by adenylation to form APS (adenosine 5'-phosphosulfate) thereby consuming ATP. The APS is then reduced by the enzyme APS reductase to form sulfite (SO₃) and AMP. In organisms that use carbon compounds as electron donors, the ATP consumed is accounted for by fermentation of the carbon substrate. The hydrogen produced during fermentation is actually what drives respiration during sulfate reduction.

ACETOGENESIS – CARBON DIOXIDE AS ELECTRON ACCEPTOR

Acetogenesis is a type of microbial metabolism that uses hydrogen (H₂) as an electron donor and carbon dioxide (CO₂) as an electron acceptor to produce acetate, the same electron donors and acceptors used in methanogenesis.

Bacteria that can autotrophically synthesize acetate are called homoacetogens. Carbon dioxide reduction in all homoacetogens occurs by the acetyl-CoA pathway. This pathway is also used for carbon fixation by autotrophic sulfate-reducing bacteria and hydrogenotrophic methanogens. Often homoacetogens can also be fermentative, using the hydrogen and carbon dioxide produced as a result of fermentation to produce acetate, which is secreted as an end product.

OTHER INORGANIC ELECTRON ACCEPTORS

Ferric iron (Fe) is a widespread anaerobic terminal electron acceptor both for autotrophic and heterotrophic organisms. Electron flow in these organisms is similar to those in electron transport, ending in oxygen or nitrate, except that in ferric iron-reducing organisms the final enzyme in this system is a ferric iron reductase. Model organisms include *Shewanella putrefaciens* and *Geobacter metallireducens*. Since some ferric iron-reducing bacteria (e.g. *G. metallireducens*) can use toxic hydrocarbons such as toluene as a carbon source, there is significant interest in using these organisms as bioremediation agents in ferric iron-rich contaminated aquifers.

Although ferric iron is the most prevalent inorganic electron acceptor, a number of organisms (including the iron-reducing bacteria mentioned above) can use other inorganic ions in anaerobic respiration. While these processes may often be less significant ecologically, they are of considerable interest for bioremediation, especially when heavy metals or radionuclides are used as electron acceptors. Examples include:

- Manganic ion (Mn) reduction to manganous ion (Mn)
- Selenate (SeO₄) reduction to selenite (SeO₃) and selenite reduction to inorganic selenium (Se)
- Arsenate (AsO₄) reduction to arsenite (AsO₃)
- Uranyl ion (UO₂²⁺) reduction to uranium dioxide (UO₂)

ORGANIC TERMINAL ELECTRON ACCEPTORS

A number of organisms, instead of using inorganic compounds as terminal electron acceptors, are able to use organic compounds to accept electrons from respiration. Examples include:

- Fumarate reduction to succinate
- Trimethylamine N-oxide (TMAO) reduction to trimethylamine (TMA)
- Dimethyl sulfoxide (DMSO) reduction to Dimethyl sulfide (DMS)
- Reductive dechlorination

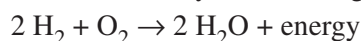
TMAO is a chemical commonly produced by fish, and when reduced to TMA produces a strong odour. DMSO is a common marine and freshwater chemical which is also odiferous when reduced to DMS. Reductive dechlorination is the process by which chlorinated organic compounds are reduced to form their non-chlorinated endproducts. As chlorinated organic compounds are often important (and difficult to degrade) environmental pollutants, reductive dechlorination is an important process in bioremediation.

CHEMOLITHOTROPHY

Chemolithotrophy is a type of metabolism where energy is obtained from the oxidation of inorganic compounds. Most chemolithotrophic organisms are also autotrophic. There are two major objectives to chemolithotrophy: the generation of energy (ATP) and the generation of reducing power (NADH).

HYDROGEN OXIDATION

Many organisms are capable of using hydrogen (H₂) as a source of energy. While several mechanisms of anaerobic hydrogen oxidation have been mentioned previously (e.g. sulfate reducing- and acetogenic bacteria), hydrogen can also be used as an energy source aerobically in the Knallgas reaction:



In these organisms, hydrogen is oxidized by a membrane-bound hydrogenase causing proton pumping via electron transfer to various quinones and cytochromes. In many organisms, a second cytoplasmic hydrogenase is used to generate reducing power in the form of NADH, which is subsequently used to fix carbon dioxide via the Calvin cycle. Hydrogen-oxidizing organisms, such as *Cupriavidus necator* (formerly *Ralstonia eutropha*), often inhabit oxic-anoxic interfaces in nature to take advantage of the hydrogen produced by anaerobic fermentative organisms while still maintaining a supply of oxygen.

SULFUR OXIDATION

Sulfur oxidation involves the oxidation of reduced sulfur compounds (such as sulfide H_2S), inorganic sulfur (S), and thiosulfate (S_2O_3) to form sulfuric acid (H_2SO_4). A classic example of a sulfur-oxidizing bacterium is *Beggiatoa*, a microbe originally described by Sergei Winogradsky, one of the founders of environmental microbiology. Another example is *Paracoccus*. Generally, the oxidation of sulfide occurs in stages, with inorganic sulfur being stored either inside or outside of the cell until needed. This two step process occurs because energetically sulfide is a better electron donor than inorganic sulfur or thiosulfate, allowing for a greater number of protons to be translocated across the membrane. Sulfur-oxidizing organisms generate reducing power for carbon dioxide fixation via the Calvin cycle using reverse electron flow, an energy-requiring process that pushes the electrons against their thermodynamic gradient to produce NADH. Biochemically, reduced sulfur compounds are converted to sulfite (SO_3) and subsequently converted to sulfate (SO_4) by the enzyme sulfite oxidase. Some organisms, however, accomplish the same oxidation using a reversal of the APS reductase system used by sulfate-reducing bacteria. In all cases the energy liberated is transferred to the electron transport chain for ATP and NADH production. In addition to aerobic sulfur oxidation, some organisms (e.g. *Thiobacillus denitrificans*) use nitrate (NO_3) as a terminal electron acceptor and therefore grow anaerobically.

FERROUS IRON (FE) OXIDATION

Ferrous iron is a soluble form of iron that is stable at extremely low pHs or under anaerobic conditions. Under aerobic, moderate pH conditions ferrous iron is oxidized spontaneously to the ferric (Fe) form and is hydrolyzed abiotically to insoluble ferric hydroxide ($\text{Fe}(\text{OH})_3$). There are three distinct types of ferrous iron-oxidizing microbes. The first are acidophiles, such as the bacteria *Acidithiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*, as well as the archaeon *Ferroplasma*. These microbes oxidize iron in environments that have a very low pH and are important in acid mine drainage. The second type of microbes oxidize ferrous iron at near-neutral pH. These micro-organisms (for example *Gallionella ferruginea*, *Leptothrix ochracea*, or *Mariprofundus ferrooxydans*) live at the oxic-anoxic interfaces and are microaerophiles. The third type of iron-oxidizing microbes are anaerobic photosynthetic bacteria such as *Rhodospseudomonas*, which use ferrous iron to produce NADH for autotrophic carbon dioxide fixation. Biochemically, aerobic iron oxidation is a very energetically poor process which therefore requires large amounts of iron to be oxidized by the enzyme rusticyanin to facilitate the formation of proton motive force. Like sulfur oxidation, reverse electron flow must be used to form the NADH used for carbon dioxide fixation via the Calvin cycle.

NITRIFICATION

Nitrification is the process by which ammonia (NH_3) is converted to nitrate (NO_3). Nitrification is actually the net result of two distinct processes: oxidation of ammonia to nitrite (NO_2) by nitrosifying bacteria (e.g. *Nitrosomonas*) and oxidation of nitrite to nitrate by the nitrite-oxidizing bacteria (e.g. *Nitrobacter*). Both of these processes are extremely energetically poor leading to very slow growth rates for both types of organisms. Biochemically, ammonia oxidation occurs by the stepwise oxidation of ammonia to hydroxylamine (NH_2OH) by the enzyme ammonia monooxygenase in the cytoplasm, followed by the oxidation of hydroxylamine to nitrite by the enzyme hydroxylamine oxidoreductase in the periplasm.

Electron and proton cycling are very complex but as a net result only one proton is translocated across the membrane per molecule of ammonia oxidized. Nitrite oxidation is much simpler, with nitrite being oxidized by the enzyme nitrite oxidoreductase coupled to proton translocation by a very short electron transport chain, again leading to very low growth rates for these organisms. Oxygen is required in both ammonia and nitrite oxidation, meaning that both nitrosifying and nitrite-oxidizing bacteria are aerobes. As in sulfur and iron oxidation,

NADH for carbon dioxide fixation using the Calvin cycle is generated by reverse electron flow, thereby placing a further metabolic burden on an already energy-poor process. In 2015, two groups independently showed the microbial genus *Nitrospira* is capable of complete nitrification (Comammox).

ANAMMOX

Anammox stands for anaerobic ammonia oxidation and the organisms responsible were relatively recently discovered, in the late 1990s. This form of metabolism occurs in members of the Planctomycetes (*e.g.* *Candidatus Brocadia anammoxidans*) and involves the coupling of ammonia oxidation to nitrite reduction. As oxygen is not required for this process, these organisms are strict anaerobes. Amazingly, hydrazine (N_2H_4 – rocket fuel) is produced as an intermediate during anammox metabolism. To deal with the high toxicity of hydrazine, anammox bacteria contain a hydrazine-containing intracellular organelle called the anammoxosome, surrounded by highly compact (and unusual) ladderane lipid membrane. These lipids are unique in nature, as is the use of hydrazine as a metabolic intermediate. Anammox organisms are autotrophs although the mechanism for carbon dioxide fixation is unclear. Because of this property, these organisms could be used to remove nitrogen in industrial wastewater treatment processes. Anammox has also been shown have widespread occurrence in anaerobic aquatic systems and has been speculated to account for approximately 50 per cent of nitrogen gas production in the ocean.

PHOTOTROPHY

Many microbes (phototrophs) are capable of using light as a source of energy to produce ATP and organic compounds such as carbohydrates, lipids, and proteins. Of these, algae are particularly significant because they are oxygenic, using water as an electron donor for electron transfer during photosynthesis. Phototrophic bacteria are found in the phyla Cyanobacteria, Chlorobi, Proteobacteria, Chloroflexi, and Firmicutes. Along with plants these microbes are responsible for all biological generation of oxygen gas on Earth. Because chloroplasts were derived from a lineage of the Cyanobacteria, the general principles of metabolism in these endosymbionts can also be applied to chloroplasts. In addition to oxygenic photosynthesis, many bacteria can also photosynthesize anaerobically, typically using sulfide (H_2S) as an electron donor to produce sulfate. Inorganic sulfur (S_0), thiosulfate (S_2O_3) and ferrous iron (Fe) can also be used by some organisms. Phylogenetically, all oxygenic photosynthetic bacteria are Cyanobacteria, while anoxygenic photosynthetic bacteria belong to the purple bacteria (Proteobacteria), Green sulfur bacteria (*e.g.* *Chlorobium*), Green non-sulfur bacteria (*e.g.* *Chloroflexus*), or the heliobacteria (Low per cent G+C Gram positives). In addition to these organisms, some microbes (*e.g.* the Archaeon *Halobacterium* or the bacterium *Roseobacter*, among others) can utilize light to produce energy using the enzyme bacteriorhodopsin, a light-driven proton pump. However, there are no known Archaea that carry out photosynthesis.

As befits the large diversity of photosynthetic bacteria, there are many different mechanisms by which light is converted into energy for metabolism. All photosynthetic organisms locate their photosynthetic reaction centers within a membrane, which may be invaginations of the cytoplasmic membrane (Proteobacteria), thylakoid membranes (Cyanobacteria), specialized antenna structures called chlorosomes (Green sulfur and non-sulfur bacteria), or the cytoplasmic membrane itself (heliobacteria). Different photosynthetic bacteria also contain different photosynthetic pigments, such as chlorophylls and carotenoids, allowing them to take advantage of different portions of the electromagnetic spectrum and thereby inhabit different niches. Some groups of organisms contain more specialized light-harvesting structures (*e.g.* phycobilisomes in Cyanobacteria and chlorosomes in Green sulfur and non-sulfur bacteria), allowing for increased efficiency in light utilization. Biochemically, anoxygenic photosynthesis is very different from oxygenic photosynthesis. Cyanobacteria (and by extension, chloroplasts) use the Z scheme of electron flow in which electrons eventually are used to form NADH. Two

different reaction centers (photosystems) are used and proton motive force is generated both by using cyclic electron flow and the quinone pool. In anoxygenic photosynthetic bacteria, electron flow is cyclic, with all electrons used in photosynthesis eventually being transferred back to the single reaction center. A proton motive force is generated using only the quinone pool. In heliobacteria, Green sulfur, and Green non-sulfur bacteria, NADH is formed using the protein ferredoxin, an energetically favorable reaction. In purple bacteria, NADH is formed by reverse electron flow due to the lower chemical potential of this reaction center. In all cases, however, a proton motive force is generated and used to drive ATP production via an ATPase. Most photosynthetic microbes are autotrophic, fixing carbon dioxide via the Calvin cycle. Some photosynthetic bacteria (*e.g.* Chloroflexus) are photoheterotrophs, meaning that they use organic carbon compounds as a carbon source for growth. Some photosynthetic organisms also fix nitrogen.

NITROGEN FIXATION

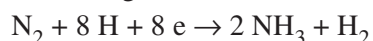
Nitrogen fixation is a process by which nitrogen in the Earth's atmosphere is converted into ammonia (NH₃) or other molecules available to living organisms. Atmospheric nitrogen or molecular dinitrogen (N₂) is relatively inert: it does not easily react with other chemicals to form new compounds. The fixation process frees nitrogen atoms from their triply bonded diatomic form, N≡N, to be used in other ways. Nitrogen fixation is essential for some forms of life because inorganic nitrogen compounds are required for the biosynthesis of the basic building blocks of plants, animals and other life forms, *e.g.*, nucleotides for DNA and RNA, the coenzyme nicotinamide adenine dinucleotide for its role in metabolism (transferring electrons between molecules), and amino acids for proteins. Therefore, as part of the nitrogen cycle, it is essential for agriculture and the manufacture of fertilizer. It is also, indirectly, relevant to the manufacture of all chemical compounds that contain nitrogen, which includes explosives, most pharmaceuticals, and dyes. Nitrogen fixation is carried out naturally in the soil by a wide range of nitrogen fixing Bacteria and Archaea, including *Azotobacter*. Some nitrogen-fixing bacteria have symbiotic relationships with some plant groups, especially legumes. Looser relationships between nitrogen-fixing bacteria and plants are often referred to as associative or non-symbiotic, as seen in nitrogen fixation occurring on rice roots. It also occurs naturally in the air by means of NO_x production by lightning. All biological nitrogen fixation is done by way of metalloenzymes called nitrogenases. These enzymes contain iron, often with a second metal, usually molybdenum but sometimes vanadium. Microorganisms that can fix nitrogen are prokaryotes (both bacteria and archaea, distributed throughout their respective domains) called diazotrophs. Some higher plants, and some animals (termites), have formed associations (symbiosis) with diazotrophs.

NON-BIOLOGICAL NATURAL NITROGEN FIXATION

Nitrogen in air can be fixed by lightning converting nitrogen and oxygen into NO_x (nitrogen oxides). NO_x may react with water to make nitrous acid or nitric acid, which seeps into the soil, where it makes nitrate, which is of use to growing plants.

BIOLOGICAL NITROGEN FIXATION

Biological nitrogen fixation was discovered by the German agronomist Hermann Hellriegel and Dutch microbiologist Martinus Beijerinck. Biological nitrogen fixation (BNF) occurs when atmospheric nitrogen is converted to ammonia by an enzyme called a nitrogenase. The overall reaction for BNF is:



The process is coupled to the hydrolysis of 16 equivalents of ATP and is accompanied by the co-formation of one molecule of H₂. The conversion of N₂ into ammonia occurs at a cluster called FeMoco, an abbreviation for the iron-molybdenum cofactor. The mechanism proceeds via a series of protonation and reduction steps wherein the FeMoco active site hydrogenates the N₂ substrate.

In free-living diazotrophs, the nitrogenase-generated ammonium is assimilated into glutamate through the glutamine synthetase/glutamate synthase pathway. The microbial genes required for nitrogen fixation are widely distributed in diverse environments. Enzymes responsible for nitrogenase action are often very susceptible to destruction by oxygen. For this reason, many bacteria cease production of the enzyme in the presence of oxygen. Many nitrogen-fixing organisms exist only in anaerobic conditions, respiring to draw down oxygen levels, or binding the oxygen with a protein such as leghemoglobin.

Microorganisms that Fix Nitrogen

Azotobacteraceae, rhizobia and *Frankia*. Several obligately anaerobic bacteria fix nitrogen including many (but not all) *Clostridium* spp. Some Archaea also fix nitrogen, including several methanogenic taxa, which are significant contributors to nitrogen fixation in anoxic soils. Cyanobacteria inhabit nearly all illuminated environments on Earth and play key roles in the carbon and nitrogen cycle of the biosphere. In general, cyanobacteria can use various inorganic and organic sources of combined nitrogen, like nitrate, nitrite, ammonium, urea, or some amino acids. Several cyanobacterial strains are also capable of diazotrophic growth, an ability that may have been present in their last common ancestor in the Archean eon. Nitrogen fixation by cyanobacteria in coral reefs can fix twice as much nitrogen as on land—around 1.8 kg of nitrogen is fixed per hectare per day (around 660 kg/ha/year). The colonial marine cyanobacterium *Trichodesmium* is thought to fix nitrogen on such a scale that it accounts for almost half of the nitrogen fixation in marine systems globally.

Root Nodule Symbioses

The Legume Family

Plants that contribute to nitrogen fixation include those of the legume family – Fabaceae – with taxa such as kudzu, clovers, soybeans, alfalfa, lupines, peanuts, and rooibos. They contain symbiotic bacteria called rhizobia within nodules in their root systems, producing nitrogen compounds that help the plant to grow and compete with other plants. When the plant dies, the fixed nitrogen is released, making it available to other plants; this helps to fertilize the soil. The great majority of legumes have this association, but a few genera (*e.g.*, *Styphnolobium*) do not. In many traditional and organic farming practices, fields are rotated through various types of crops, which usually include one consisting mainly or entirely of clover or buckwheat (non-legume family Polygonaceae), often referred to as “green manure”.

The efficiency of nitrogen fixation in soil is dependent on many factors, including the legume as well as air and soil conditions. For example, nitrogen fixation by red clover can range from 50 - 200 lb/acre depending on these variables. Inga alley farming relies on the leguminous genus *Inga*, a small tropical, tough-leaved, nitrogen-fixing tree.

Non-leguminous

Although by far the majority of plant species able to form nitrogen-fixing root nodules are in the legume family Fabaceae, there are exceptions:

- *Parasponia*, a tropical genus in the Cannabaceae also able to interact with rhizobia and form nitrogen-fixing nodules
- Actinorhizal plants such as alder and bayberry can also form nitrogen-fixing nodules, thanks to a symbiotic association with *Frankia* bacteria. These plants belong to 25 genera distributed among 8 plant families.

The ability to fix nitrogen is present in the families listed below. They belong to the orders Cucurbitales, Fagales, and Rosales, which together with the Fabales form a clade of eurosids. The ability to fix nitrogen is not universally present in these families. For example, of 122 genera in the Rosaceae, only 4 genera are capable of fixing nitrogen. Fabales were the first lineage to branch off this nitrogen-fixing clade; thus, the ability to fix nitrogen may be plesiomorphic and subsequently lost in most descendants of the original nitrogen-fixing plant; however, it may be that the basic genetic and physiological requirements were present in an incipient state in the last common ancestors of all these plants, but only evolved to full function in some of them.

Family: Genera	...	Coriariaceae:	...	Myricaceae:	...	Rhamnaceae:	...	Rosaceae:
Betulaceae:	...	Coriaria	...	Comptonia	...	Ceanothus	...	Cercocarpus (mountain mahoganies)
Alnus (alders)		Datisceae:		(sweetfern)		Colletia		Chamaebatia (mountain miseries)
Cannabaceae:		Datisca		Morella		Discaria		Dryas
Trema		Elaeagnaceae:		Myrica		Kentrothamnus		Purshia/Cowania
Casuarinaceae:		Elaeagnus		(bayberries)		Retanilla		(bitterbrushes/cliffroses)
Allocasuarina		(silverberries)				Talguenea		
Casuarina		Hippophae (sea-buckthorns)				Trevoa		
Ceuthostoma		Shepherdia						
Gymnostoma		(buffaloberries)						

There are also several nitrogen-fixing symbiotic associations that involve cyanobacteria (such as Nostoc):

- Some lichens such as *Lobaria* and *Peltigera*
- Mosquito fern (*Azolla* species)
- Cycads
- *Gunnera*

Endosymbiosis in Diatoms

Rhopalodia gibba, a diatom alga, is a eukaryote with cyanobacterial N_2 -fixing endosymbiont organelles. The spheroid bodies reside in the cytoplasm of the diatoms and are inseparable from their hosts.

INDUSTRIAL NITROGEN FIXATION

The possibility that atmospheric nitrogen reacts with certain chemicals was first observed by Desfosses in 1828. He observed that mixtures of alkali metal oxides and carbon react at high temperatures with nitrogen. With the use of barium carbonate as starting material the first commercially used process became available in the 1860s developed by Margueritte and Sourdeval. The resulting barium cyanide could be reacted with steam yielding ammonia. In 1898 Adolph Frank and Nikodem Caro decoupled the process and first produced calcium carbide and in a subsequent step reacted it with nitrogen to calcium cyanamide. The Ostwald process for the production of nitric acid was discovered in 1902. Frank-Caro process and Ostwald process dominated the industrial fixation of nitrogen until the discovery of the Haber process in 1909. Prior to 1900, Nikola Tesla also experimented with the industrial fixation of nitrogen “by using currents of extremely high frequency or rate of vibration”.

Haber Process

Artificial fertilizer production is now the largest source of human-produced fixed nitrogen in the Earth's ecosystem. Ammonia is a required precursor to fertilizers, explosives, and other products. The most common method is the Haber process. The Haber process requires high pressures (around 200 atm) and high temperatures (at least 400 °C), routine conditions for industrial catalysis. This highly efficient process uses natural gas as a hydrogen source and air as a nitrogen source.

Much research has been conducted on the discovery of catalysts for nitrogen fixation, often with the goal of reducing the energy required for this conversion. However, such research has thus far failed to even approach the efficiency and ease of the Haber process. Many compounds react with atmospheric nitrogen to give dinitrogen complexes. The first dinitrogen complex to be reported was $\text{Ru}(\text{NH}_3)_5(\text{N}_2)$.

Ambient Nitrogen Reduction

Catalytic chemical nitrogen fixation at ambient conditions is an ongoing scientific endeavor. Guided by the example of nitrogenase, this area of homogeneous catalysis is ongoing, with particular emphasis on hydrogenation to give ammonia.

Metallic lithium has long been known for burning in an atmosphere of nitrogen and then converting to lithium nitride. Hydrolysis of the resulting nitride gives ammonia. In a related process, trimethylsilyl chloride, lithium, and nitrogen react in the presence of a catalyst to give tris(trimethylsilyl)amine. Tris(trimethylsilyl)amine can then be used for reaction with α,δ,ω -triketones to give tricyclic pyrroles. Processes involving lithium metal are however of no practical interest since they are noncatalytic and re-reducing the Li ion residue is difficult. Beginning in the 1960s several homogeneous systems were identified that convert nitrogen to ammonia, sometimes even catalytically but often operating via ill-defined mechanisms. The original discovery is described in an early review:

- “Vol’pin and co-workers, using a non-protic Lewis acid, aluminium tribromide, were able to demonstrate the truly catalytic effect of titanium by treating dinitrogen with a mixture of titanium tetrachloride, metallic aluminium, and aluminium tribromide at 50 °C, either in the absence or in the presence of a solvent, *e.g.* benzene. As much as 200 mol of ammonia per mol of TiCl_4 was obtained after hydrolysis....”

The quest for well defined intermediates led to the characterization of many transition metal dinitrogen complexes. Few of these well defined complexes function catalytically, their behaviour illuminated likely stages in nitrogen fixation. Most fruitful of all of these early studies focused on $\text{M}(\text{N}_2)_2(\text{dppe})_2$ ($\text{M} = \text{Mo}, \text{W}$). For example, double protonation of such low valent complexes gave intermediates with the linkage $\text{M}=\text{N}-\text{NH}_2$. In 1995, a molybdenum(III) amido complex was discovered that cleaved N_2 to give the corresponding molybdenum(VI) nitride. This and related terminal nitrido complexes have been used to make nitriles.

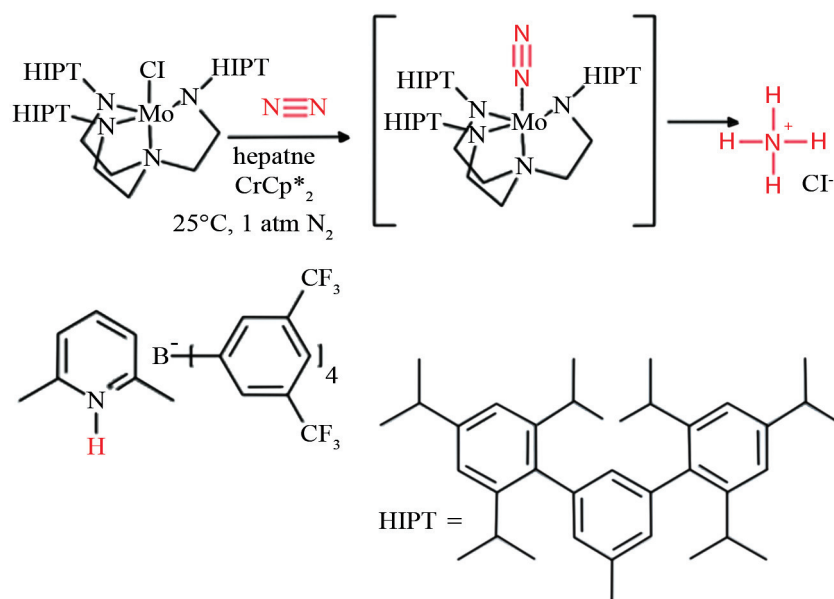


Fig. Synthetic nitrogen reduction.

In 2003 a related molybdenum amido complex was found to catalyze the reduction of N_2 . In addition to a source of protons, the catalyst requires a strong reducing agent. However, this catalytic reduction fixates only a few nitrogen molecules. In these systems, like the biological one, hydrogen is provided to the substrate heterolytically, by means of protons and reducing equivalents rather than with H_2 itself. In 2011 Arashiba et al. reported yet another system with a catalyst again based on molybdenum but with a diphosphorus pincer ligand. Photolytic nitrogen splitting is also considered.

Genetics

Genetics is the study of genes, genetic variation, and heredity in living organisms. It is generally considered a field of biology, but intersects frequently with many other life sciences and is strongly linked with the study of information systems. The discoverer of genetics is Gregor Mendel, a late 19th-century scientist and Augustinian friar. Mendel studied “trait inheritance”, patterns in the way traits are handed down from parents to offspring. He observed that organisms (pea plants) inherit traits by way of discrete “units of inheritance”. This term, still used today, is a somewhat ambiguous definition of what is referred to as a gene. Trait inheritance and molecular inheritance mechanisms of genes are still primary principles of genetics in the 21st century, but modern genetics has expanded beyond inheritance to studying the function and behaviour of genes. Gene structure and function, variation, and distribution are studied within the context of the cell, the organism (*e.g.* dominance), and within the context of a population. Genetics has given rise to a number of subfields, including epigenetics and population genetics.

Organisms studied within the broad field span the domains of life (archaea, bacteria, and eukarya). Genetic processes work in combination with an organism’s environment and experiences to influence development and behaviour, often referred to as nature versus nurture. The intracellular or extracellular environment of a cell or organism may switch gene transcription on or off. A classic example is two seeds of genetically identical corn, one placed in a temperate climate and one in an arid climate. While the average height of the two corn stalks may be genetically determined to be equal, the one in the arid climate only grows to half the height of the one in the temperate climate due to lack of water and nutrients in its environment.

HISTORY

The history of genetics dates from the classical era with contributions by Hippocrates, Aristotle and Epicurus. Modern genetics began with the work of the Augustinian friar Gregor Johann Mendel. His work on pea plants, published in 1866, described what came to be known as Mendelian inheritance. Many theories of heredity proliferated in the centuries before and for several decades after Mendel’s work.

The year 1900 marked the “rediscovery of Mendel” by Hugo de Vries, Carl Correns and Erich von Tschermak, and by 1915 the basic principles of Mendelian genetics had been applied to a wide variety of organisms—most notably the fruit fly *Drosophila melanogaster*. Led by Thomas Hunt Morgan and his fellow “drosophilists”, geneticists developed the Mendelian model, which was widely accepted by 1925. Alongside experimental work, mathematicians developed the statistical framework of population genetics, bringing genetic explanations into the study of evolution. With the basic patterns of genetic inheritance established, many biologists turned to investigations of the physical nature of the gene. In the 1940s and early 1950s, experiments pointed to DNA as the portion of chromosomes (and perhaps other nucleoproteins) that held genes. A focus on new model organisms such as viruses and bacteria, along with the discovery of the double helical structure of DNA in 1953, marked the transition to the era of molecular genetics. In the following years, chemists developed techniques for

sequencing both nucleic acids and proteins, while others worked out the relationship between the two forms of biological molecules: the genetic code. The regulation of gene expression became a central issue in the 1960s; by the 1970s gene expression could be controlled and manipulated through genetic engineering. In the last decades of the 20th century, many biologists focused on large-scale genetics projects, sequencing entire genomes.

PRE-MENDELIAN IDEAS ON HEREDITY

Ancient Theories

The most influential early theories of heredity were that of Hippocrates and Aristotle. Hippocrates' theory (possibly based on the teachings of Anaxagoras) was similar to Darwin's later ideas on pangenesis, involving heredity material that collects from throughout the body. Aristotle suggested instead that the (nonphysical) form-giving principle of an organism was transmitted through semen (which he considered to be a purified form of blood) and the mother's menstrual blood, which interacted in the womb to direct an organism's early development. For both Hippocrates and Aristotle—and nearly all Western scholars through to the late 19th century—the inheritance of acquired characters was a supposedly well-established fact that any adequate theory of heredity had to explain. At the same time, individual species were taken to have a fixed essence; such inherited changes were merely superficial. The Athenian philosopher Epicurus observed families and proposed the contribution of both males and females of hereditary characters ("sperm atoms"), noticed dominant and recessive types of inheritance and described segregation and independent assortment of "sperm atoms"

In the Charaka Samhita of 300CE, ancient Indian medical writers saw the characteristics of the child as determined by four factors:

- Those from the mother's reproductive material,
- Those from the father's sperm,
- Those from the diet of the pregnant mother and
- Those accompanying the soul which enters into the foetus. Each of these four factors had four parts creating sixteen factors of which the karma of the parents and the soul determined which attributes predominated and thereby gave the child its characteristics.

In the 9th century CE, the Afro-Arab writer Al-Jahiz considered the effects of the environment on the likelihood of an animal to survive. In 1000 CE, the Arab physician, Abu al-Qasim al-Zahrawi (known as Albucasis in the West) was the first physician to describe clearly the hereditary nature of haemophilia in his *Al-Tasrif*. In 1140 CE, Judah HaLevi described dominant and recessive genetic traits in *The Kuzari*.

Plant Systematics and Hybridization

In the 18th century, with increased knowledge of plant and animal diversity and the accompanying increased focus on taxonomy, new ideas about heredity began to appear. Linnaeus and others (among them Joseph Gottlieb Kölreuter, Carl Friedrich von Gärtner, and Charles Naudin) conducted extensive experiments with hybridization, especially species hybrids. Species hybridizers described a wide variety of inheritance phenomena, include hybrid sterility and the high variability of back-crosses. Plant breeders were also developing an array of stable varieties in many important plant species. In the early 19th century, Augustin Sageret established the concept of dominance, recognizing that when some plant varieties are crossed, certain characteristics (present in one parent) usually appear in the offspring; he also found that some ancestral characteristics found in neither parent may appear in offspring. However, plant breeders made little attempt to establish a theoretical foundation for their work or to share their knowledge with current work of physiology, although Gartons Agricultural Plant Breeders in England explained their system.

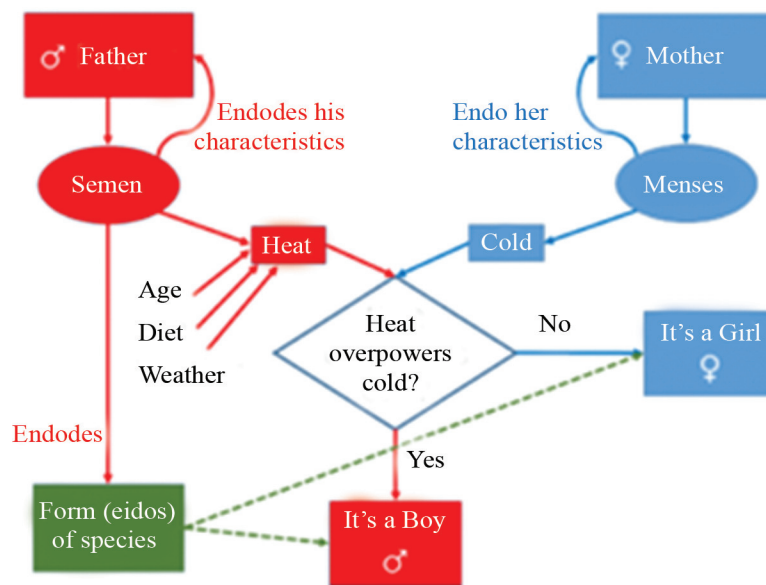


Fig. Aristotle's model of transmission of movements from parents to child, and of form from the father. The model is not fully symmetric.

MENDEL

In breeding experiments between 1856 and 1865, Gregor Mendel first traced inheritance patterns of certain traits in pea plants and showed that they obeyed simple statistical rules with some traits being dominant and others being recessive. These patterns of Mendelian inheritance demonstrated that application of statistics to inheritance could be highly useful; they also contradicted 19th century theories of blending inheritance as the traits remained discrete through multiple generation of hybridization. From his statistical analysis Mendel defined a concept that he described as a character (which in his mind holds also for “determinant of that character”). In only one sentence of his historical paper he used the term “factors” to designate the “material creating” the character: “So far as experience goes, we find it in every case confirmed that constant progeny can only be formed when the egg cells and the fertilizing pollen are of like character, so that both are provided with the material for creating quite similar individuals, as is the case with the normal fertilization of pure species. We must therefore regard it as certain that exactly similar factors must be at work also in the production of the constant forms in the hybrid plants.” (Mendel, 1866). Mendel’s work was published in 1866 as “*Versuche über Pflanzen-Hybriden*” (*Experiments on Plant Hybridization*) in the *Verhandlungen des Naturforschenden Vereins zu Brünn* (*Proceedings of the Natural History Society of Brünn*), following two lectures he gave on the work in early 1866.

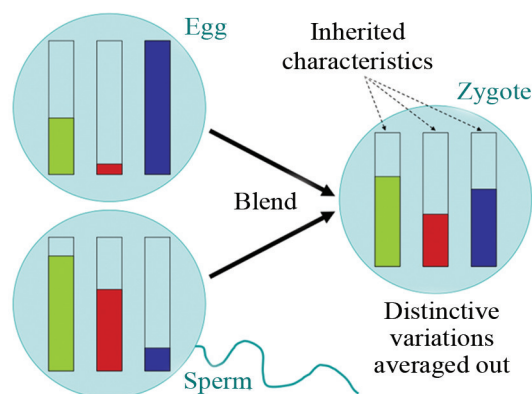


Fig. Blending inheritance leads to the averaging out of every characteristic, which would make evolution by natural selection impossible.

POST-MENDEL, PRE-REDISCOVERY

Pangenesis

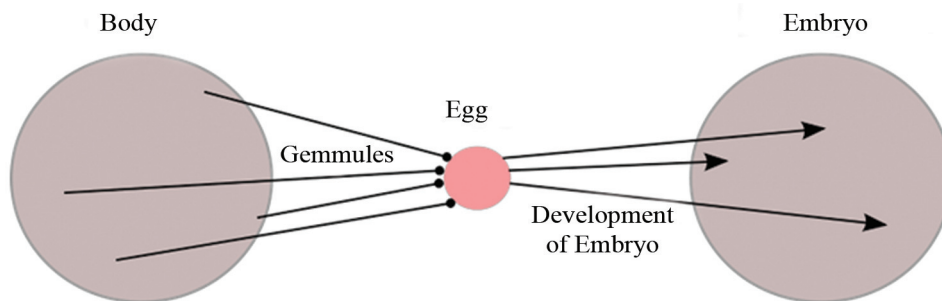


Fig. Diagram of Charles Darwin's pangenesis theory. Every part of the body emits tiny particles, gemmules, which migrate to the gonads and contribute to the fertilised egg and so to the next generation. The theory implied that changes to the body during an organism's life would be inherited, as proposed in Lamarckism.

Mendel's work was published in a relatively obscure scientific journal, and it was not given any attention in the scientific community. Instead, discussions about modes of heredity were galvanized by Darwin's theory of evolution by natural selection, in which mechanisms of non-Lamarckian heredity seemed to be required. Darwin's own theory of heredity, pangenesis, did not meet with any large degree of acceptance. A more mathematical version of pangenesis, one which dropped much of Darwin's Lamarckian holdovers, was developed as the "biometrical" school of heredity by Darwin's cousin, Francis Galton.

Germ plasm

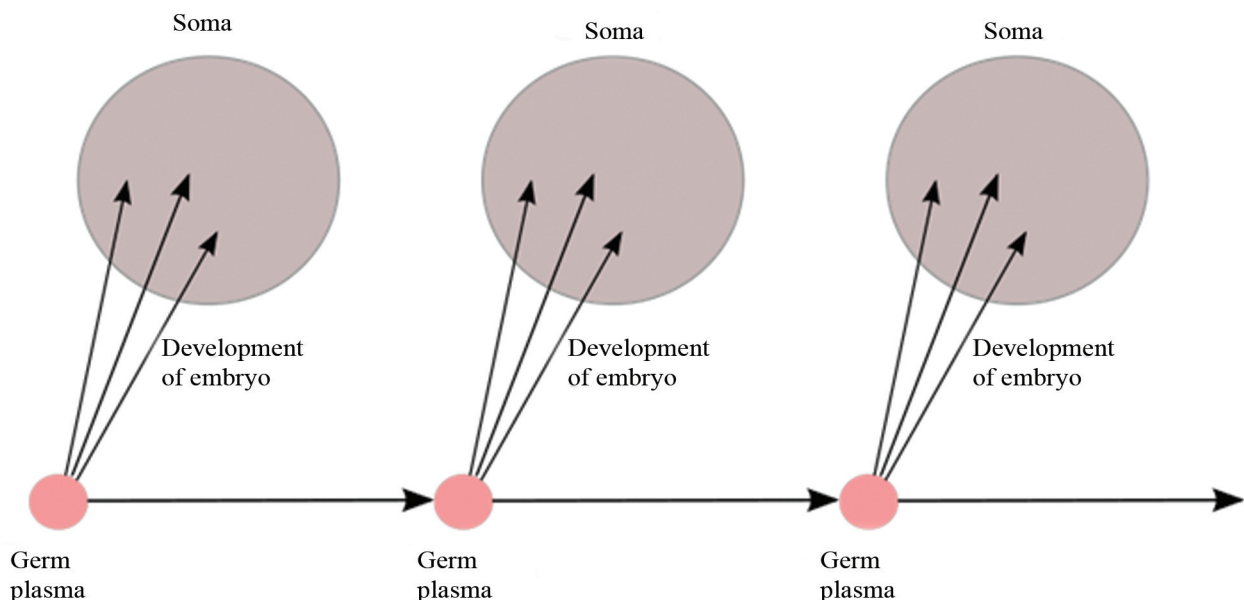


Fig. August Weismann's germ plasm theory. The hereditary material, the germ plasm, is confined to the gonads. Somatic cells (of the body) develop afresh in each generation from the germ plasm.

In 1883 August Weismann conducted experiments involving breeding mice whose tails had been surgically removed. His results — that surgically removing a mouse's tail had no effect on the tail of its offspring — challenged the theories of pangenesis and Lamarckism, which held that changes to an organism during its lifetime could be inherited by its descendants. Weismann proposed the germ plasm theory of inheritance, which held that hereditary information was carried only in sperm and egg cells.

REDISCOVERY OF MENDEL

Hugo de Vries wondered what the nature of germ plasm might be, and in particular he wondered whether or not germ plasm was mixed like paint or whether the information was carried in discrete packets that remained unbroken. In the 1890s he was conducting breeding experiments with a variety of plant species and in 1897 he published a paper on his results that stated that each inherited trait was governed by two discrete particles of information, one from each parent, and that these particles were passed along intact to the next generation. In 1900 he was preparing another paper on his further results when he was shown a copy of Mendel's 1866 paper by a friend who thought it might be relevant to de Vries's work. He went ahead and published his 1900 paper without mentioning Mendel's priority. Later that same year another botanist, Carl Correns, who had been conducting hybridization experiments with maize and peas, was searching the literature for related experiments prior to publishing his own results when he came across Mendel's paper, which had results similar to his own. Correns accused de Vries of appropriating terminology from Mendel's paper without crediting him or recognizing his priority. At the same time another botanist, Erich von Tschermak was experimenting with pea breeding and producing results like Mendel's. He too discovered Mendel's paper while searching the literature for relevant work. In a subsequent paper de Vries praised Mendel and acknowledged that he had only extended his earlier work.

EMERGENCE OF MOLECULAR GENETICS

After the rediscovery of Mendel's work there was a feud between William Bateson and Pearson over the hereditary mechanism, solved by Ronald Fisher in his work "The Correlation Between Relatives on the Supposition of Mendelian Inheritance".

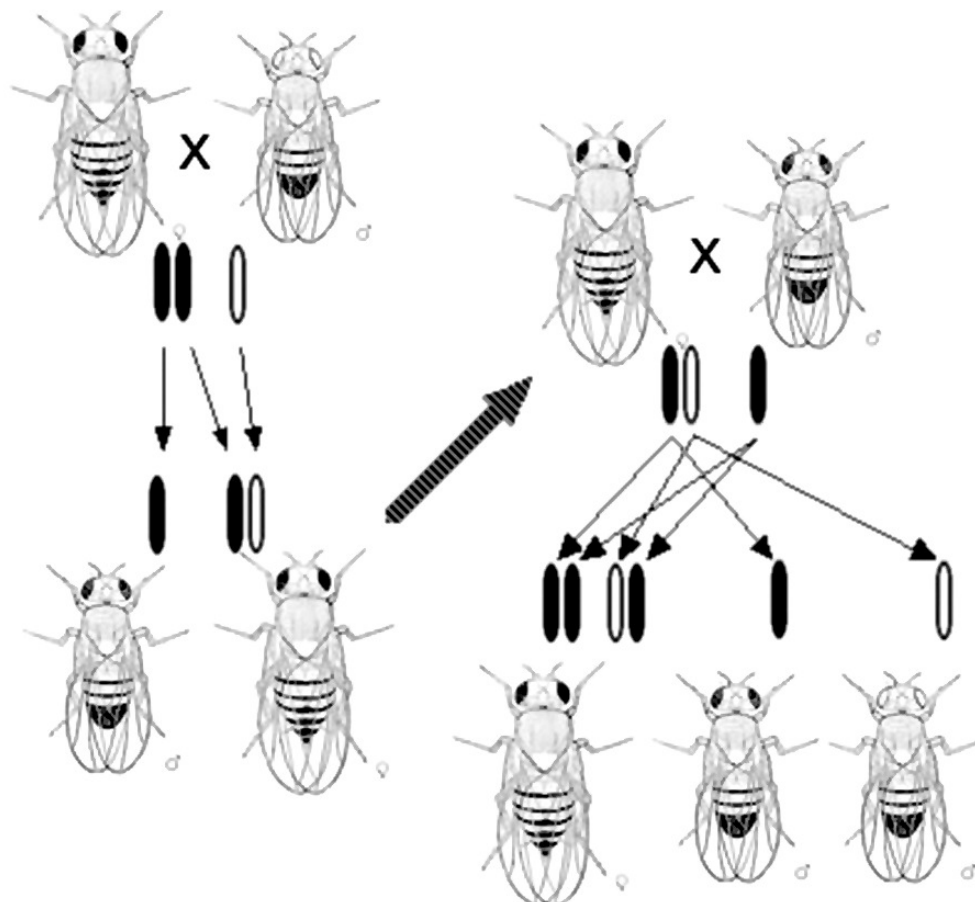


Fig. Thomas Hunt Morgan discovered sex linked inheritance of the white eyed mutation in the fruit fly *Drosophila* in 1910, implying the gene was on the sex chromosome.

In 1910, Thomas Hunt Morgan showed that genes reside on specific chromosomes. He later showed that genes occupy specific locations on the chromosome. With this knowledge, Morgan and his students began the first chromosomal map of the fruit fly *Drosophila melanogaster*. In 1928, Frederick Griffith showed that genes could be transferred. In what is now known as Griffith's experiment, injections into a mouse of a deadly strain of bacteria that had been heat-killed transferred genetic information to a safe strain of the same bacteria, killing the mouse.

A series of subsequent discoveries led to the realization decades later that the genetic material is made of DNA (deoxyribonucleic acid). In 1941, George Wells Beadle and Edward Lawrie Tatum showed that mutations in genes caused errors in specific steps in metabolic pathways. This showed that specific genes code for specific proteins, leading to the "one gene, one enzyme" hypothesis. Oswald Avery, Colin Munro MacLeod, and Maclyn McCarty showed in 1944 that DNA holds the gene's information. In 1952, Rosalind Franklin and Raymond Gosling produced a strikingly clear x-ray diffraction pattern indicating a helical form, and in 1953, James D. Watson and Francis Crick demonstrated the molecular structure of DNA. Together, these discoveries established the central dogma of molecular biology, which states that proteins are translated from RNA which is transcribed by DNA.

This dogma has since been shown to have exceptions, such as reverse transcription in retroviruses. In 1972, Walter Fiers and his team at the University of Ghent were the first to determine the sequence of a gene: the gene for bacteriophage MS2 coat protein. Richard J. Roberts and Phillip Sharp discovered in 1977 that genes can be split into segments. This led to the idea that one gene can make several proteins. The successful sequencing of many organisms' genomes has complicated the molecular definition of the gene. In particular, genes do not always sit side by side on DNA like discrete beads. Instead, regions of the DNA producing distinct proteins may overlap, so that the idea emerges that "genes are one long continuum".

It was first hypothesized in 1986 by Walter Gilbert that neither DNA nor protein would be required in such a primitive system as that of a very early stage of the earth if RNA could serve both as a catalyst and as genetic information storage processor. The modern study of genetics at the level of DNA is known as molecular genetics and the synthesis of molecular genetics with traditional Darwinian evolution is known as the modern evolutionary synthesis.

EARLY TIMELINE

- 1856-1863: Mendel studied the inheritance of traits between generations based on experiments involving garden pea plants. He deduced that there is a certain tangible essence that is passed on between generations from both parents. Mendel established the basic principles of inheritance, namely, the principles of dominance, independent assortment, and segregation.
- 1866: Austrian Augustinian monk Gregor Mendel's paper, *Experiments on Plant Hybridization*, published.
- 1869: Friedrich Miescher discovers a weak acid in the nuclei of white blood cells that today we call DNA. In 1871 he isolated cell nuclei, separated the nucleic cells from bandages and then treated them with pepsin (an enzyme which breaks down proteins). From this, he recovered an acidic substance which he called "nuclein."
- 1880-1890: Walther Flemming, Eduard Strasburger, and Edouard Van Beneden elucidate chromosome distribution during cell division
- 1889: Richard Altmann purified protein free DNA. However, the nucleic acid was not as pure as he had assumed. It was determined later to contain a large amount of protein.
- 1889: Hugo de Vries postulates that "inheritance of specific traits in organisms comes in particles", naming such particles "(pan)genes"

- 1902: Archibald Garrod discovered inborn errors of metabolism. An explanation for epistasis is an important manifestation of Garrod's research, albeit indirectly. When Garrod studied alkaptonuria, a disorder that makes urine quickly turn black due to the presence of gentesate, he noticed that it was prevalent among populations whose parents were closely related.
- 1903: Walter Sutton and Theodor Boveri independently hypothesizes that chromosomes, which segregate in a Mendelian fashion, are hereditary units. Boveri was studying sea urchins when he found that all the chromosomes in the sea urchins had to be present for proper embryonic development to take place. Sutton's work with grasshoppers showed that chromosomes occur in matched pairs of maternal and paternal chromosomes which separate during meiosis. He concluded that this could be "the physical basis of the Mendelian law of heredity."
- 1905: William Bateson coins the term "genetics" in a letter to Adam Sedgwick (Zoologist) and at a meeting in 1906
- 1908: G.H. Hardy and Wilhelm Weinberg proposed the Hardy-Weinberg equilibrium model which describes the frequencies of alleles in the gene pool of a population, which are under certain specific conditions, as constant and at a state of equilibrium from generation to generation unless specific disturbing influences are introduced.
- 1910: Thomas Hunt Morgan shows that genes reside on chromosomes while determining the nature of sex-linked traits by studying *Drosophila melanogaster*. He determined that the white-eyed mutant was sex-linked based on Mendelian's principles of segregation and independent assortment.
- 1911: Alfred Sturtevant, one of Morgan's students, invented the procedure of linkage mapping which is based on the frequency of recombination.
- 1913: Alfred Sturtevant makes the first genetic map of a chromosome
- 1913: Gene maps show chromosomes containing linear arranged genes
- 1918: Ronald Fisher publishes "The Correlation Between Relatives on the Supposition of Mendelian Inheritance" the modern synthesis of genetics and evolutionary biology starts.
- 1920: Lysenkoism Started, during Lysenkoism they stated that the hereditary factor are not only in the nucleus, but also in the cytoplasm, though they called it living protoplasm.
- 1923: Frederick Griffith studied bacterial transformation and observed that DNA carries genes responsible for pathogenicity.
- 1928: Frederick Griffith discovers that hereditary material from dead bacteria can be incorporated into live bacteria.
- 1930s–1950s: Joachim Hämmerling conducted experiments with *Acetabularia* in which he began to distinguish the contributions of the nucleus and the cytoplasm substances (later discovered to be DNA and mRNA, respectively) to cell morphogenesis and development.
- 1931: Crossing over is identified as the cause of recombination; the first cytological demonstration of this crossing over was performed by Barbara McClintock and Harriet Creighton
- 1933: Jean Brachet, while studying virgin sea urchin eggs, suggested that DNA is found in cell nucleus and that RNA is present exclusively in the cytoplasm. At the time, "yeast nucleic acid" (RNA) was thought to occur only in plants, while "thymus nucleic acid" (DNA) only in animals. The latter was thought to be a tetramer, with the function of buffering cellular pH.
- 1933: Thomas Morgan received the Nobel prize for linkage mapping. His work elucidated the role played by the chromosome in heredity.
- 1941: Edward Lawrie Tatum and George Wells Beadle show that genes code for proteins;
- 1943: Luria–Delbrück experiment: this experiment showed that genetic mutations conferring resistance to bacteriophage arise in the absence of selection, rather than being a response to selection.

THE DNA ERA

- 1944: The Avery–MacLeod–McCarty experiment isolates DNA as the genetic material (at that time called transforming principle)
- 1947: Salvador Luria discovers reactivation of irradiated phage, stimulating numerous further studies of DNA repair processes in bacteriophage, and other organisms, including humans
- 1948: Barbara McClintock discovers transposons in maize
- 1950: Erwin Chargaff determined the pairing method of nitrogenous bases. Chargaff and his team studied the DNA from multiple organisms and found three things (also known as Chargaff's rules). First, the concentration of the pyrimidines (guanine and adenine) are always found in the same amount as one another. Second, the concentration of purines (cytosine and thymidine) are also always the same. Lastly, Chargaff and his team found the proportion of pyrimidines and purines correspond each other.

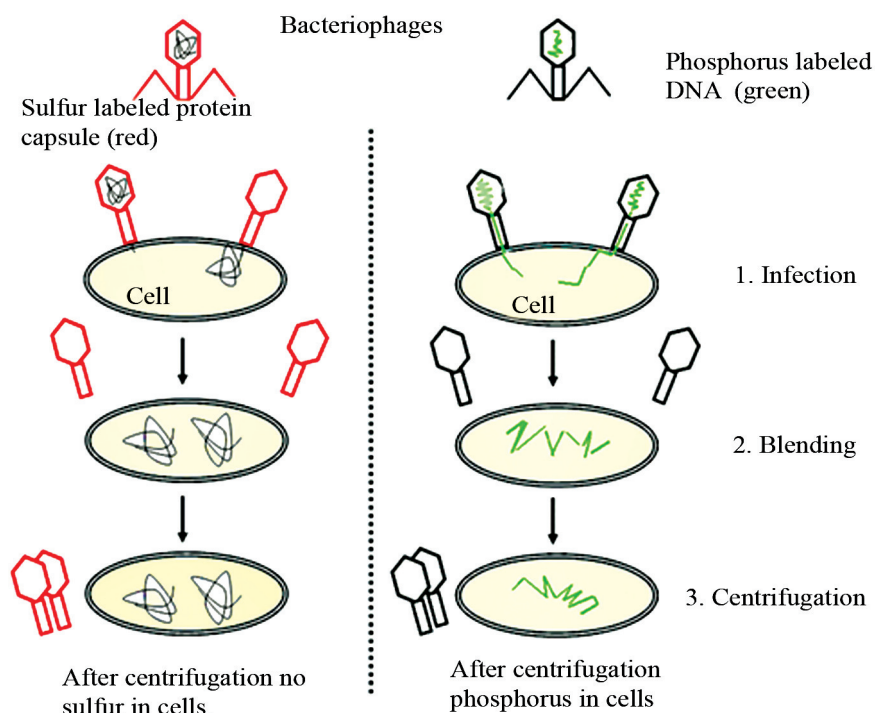


Fig. Hershey–Chase experiment proves that phage genetic material is DNA.

- 1952: The Hershey–Chase experiment proves the genetic information of phages (and, by implication, all other organisms) to be DNA.
- 1952: an X-ray diffraction image of DNA taken by Raymond Gosling in May 1952, a student supervised by Rosalind Franklin
- 1953: DNA structure is resolved to be a double helix by Rosalind Franklin, James Watson and Francis Crick
- 1955: Alexander R. Todd determined the chemical makeup of nitrogenous bases. Todd also successfully synthesized adenosine triphosphate (ATP) and flavin adenine dinucleotide (FAD). He was awarded the Nobel prize in Chemistry in 1957 for his contributions in the scientific knowledge of nucleotides and nucleotide co-enzymes.
- 1955: Joe Hin Tjio, while working in Albert Levan's lab, determined the number of chromosomes in humans to be of 46. Tjio was attempting to refine an established technique to separate chromosomes onto glass slides by conducting a study of human embryonic lung tissue, when he saw that there were 46 chromosomes rather than 48. This revolutionized the world of cytogenetics.

- 1957: Arthur Kornberg with Severo Ochoa synthesized DNA in a test tube after discovering the means by which DNA is duplicated. DNA polymerase 1 established requirements for in vitro synthesis of DNA. Kornberg and Ochoa were awarded the Nobel Prize in 1959 for this work.
- 1957/1958: Robert W. Holley, Marshall Nirenberg, Har Gobind Khorana proposed the nucleotide sequence of the tRNA molecule. Francis Crick had proposed the requirement of some kind of adapter molecule and it was soon identified by Holey, Nirenberg and Khorana. These scientists help explain the link between a messenger RNA nucleotide sequence and a polypeptide sequence. In the experiment, they purified tRNAs from yeast cells and were awarded the Nobel prize in 1968.

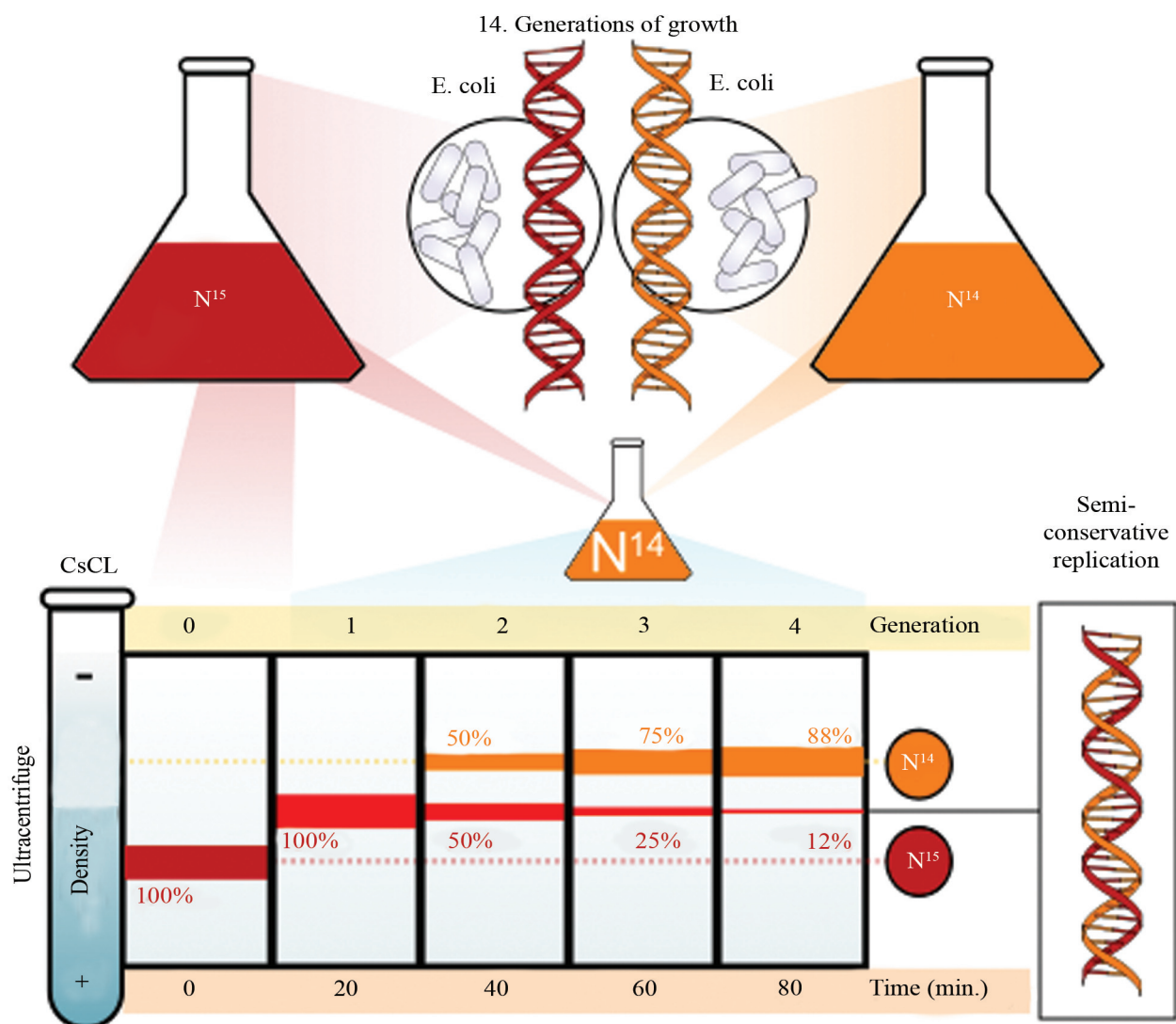


Fig. Meselson–Stahl experiment demonstrates DNA is semiconservatively replicated.

- 1958: The Meselson–Stahl experiment demonstrates that DNA is semiconservatively replicated.
- 1960: Jacob and collaborators discover the operon, a group of genes whose expression is coordinated by an operator.
- 1961: Francis Crick and Sydney Brenner discovered frame shift mutations. In the experiment, proflavin-induced mutations of the T4 bacteriophage gene (rIIB) were isolated. Proflavin causes mutations by inserting itself between DNA bases, typically resulting in insertion or deletion of a single base pair. The mutants could not produce functional rIIB protein. These mutations were used

to demonstrate that three sequential bases of the rIIB gene's DNA specify each successive amino acid of the encoded protein. Thus the genetic code is a triplet code, where each triplet (called a codon) specifies a particular amino acid.

- 1961: Sydney Brenner, Francois Jacob and Matthew Meselson identified the function of messenger RNA.
- 1961 - 1967: Combined efforts of scientists "crack" the genetic code, including Marshall Nirenberg, Har Gobind Khorana, Sydney Brenner and Francis Crick.
- 1964: Howard Temin showed using RNA viruses that the direction of DNA to RNA transcription can be reversed
- 1964: Lysenkoism Ended
- 1966: Marshall W. Nirenberg, Philip Leder, Har Gobind Khorana cracked the genetic code by using RNA homopolymer and heteropolymer experiments, through which they figured out which triplets of RNA were translated into what amino acids in yeast cells.
- 1969: Molecular hybridization of radioactive DNA to the DNA of cytological preparation. by Pardue, M. L. and Gall, J. G.
- 1970: Restriction enzymes were discovered in studies of a bacterium, *Haemophilus influenzae*, by Hamilton O. Smith and Daniel Nathans, enabling scientists to cut and paste DNA.
- 1972: Stanley Norman Cohen and Herbert Boyer at UCSF and Stanford University constructed Recombinant DNA which can be formed by using restriction Endonuclease to cleave the DNA and DNA ligase to reattach the "sticky ends" into a bacterial plasmid.

THE GENOMICS ERA

- 1972: Walter Fiers and his team were the first to determine the sequence of a gene: the gene for bacteriophage MS2 coat protein.
- 1976: Walter Fiers and his team determine the complete nucleotide-sequence of bacteriophage MS2-RNA
- 1976: Yeast genes expressed in *E. coli* for the first time.
- 1977: DNA is sequenced for the first time by Fred Sanger, Walter Gilbert, and Allan Maxam working independently. Sanger's lab sequence the entire genome of bacteriophage ϕ -X174.
- In the late 1970s: nonisotopic methods of nucleic acid labeling were developed. The subsequent improvements in the detection of reporter molecules using immunocytochemistry and immunofluorescence, in conjunction with advances in fluorescence microscopy and image analysis, have made the technique safer, faster and reliable
- 1980: Paul Berg, Walter Gilbert and Frederick Sanger developed methods of mapping the structure of DNA. In 1972, recombinant DNA molecules were produced in Paul Berg's Stanford University laboratory. Berg was awarded the 1980 Nobel Prize in Chemistry for constructing recombinant DNA molecules that contained phage lambda genes inserted into the small circular DNA mol.
- 1980: Stanley Norman Cohen and Herbert Boyer received first U.S. patent for gene cloning, by proving the successful outcome of cloning a plasmid and expressing a foreign gene in bacteria to produce a "protein foreign to a unicellular organism." These two scientist were able to replicate proteins such as HGH, Erythropoietin and Insulin. The patent earned about \$300 million in licensing royalties for Stanford.
- 1982: The U.S. Food and Drug Administration (FDA) approved the release of the first genetically engineered human insulin, originally biosynthesized using recombination DNA methods by Genentech in 1978. Once approved, the cloning process lead to mass production of humulin (under license by Eli Lilly and Co.).

- 1983: Kary Banks Mullis invents the polymerase chain reaction enabling the easy amplification of DNA
- 1983: Barbara McClintock was awarded the Nobel Prize in Physiology or Medicine for her discovery of mobile genetic elements. McClintock studied transposon-mediated mutation and chromosome breakage in maize and published her first report in 1948 on transposable elements or transposons. She found that transposons were widely observed in corn, although her ideas weren't widely granted attention until the 1960s and 1970s when the same phenomenon was discovered in bacteria and *Drosophila melanogaster*.
- 1985: Alec Jeffreys announced DNA fingerprinting method. Jeffreys was studying DNA variation and the evolution of gene families in order to understand disease causing genes. In an attempt to develop a process to isolate many mini-satellites at once using chemical probes, Jeffreys took x-ray films of the DNA for examination and noticed that mini-satellite regions differ greatly from one person to another. In a DNA fingerprinting technique, a DNA sample is digested by treatment with specific nucleases or Restriction endonuclease and then the fragments are separated by electrophoresis producing a template distinct to each individual banding pattern of the gel.
- 1986: Jeremy Nathans found genes for colour vision and colour blindness, working with David Hogness, Douglas Vollrath and Ron Davis as they were studying the complexity of the retina.
- 1987: Yoshizumi Ishino accidentally discovers and describes part of a DNA sequence which later will be called CRISPR
- 1989: Thomas Cech discovered that RNA can catalyze chemical reactions, making for one of the most important breakthroughs in molecular genetics, because it elucidates the true function of poorly understood segments of DNA.
- 1989: The human gene that encodes the CFTR protein was sequenced by Francis Collins and Lap-Chee Tsui. Defects in this gene cause cystic fibrosis.
- 1992: American and British scientists unveiled a technique for testing embryos in-vitro (Amniocentesis) for genetic abnormalities such as Cystic fibrosis and Hemophilia.
- 1993: Phillip Allen Sharp and Richard Roberts awarded the Nobel Prize for the discovery that genes in DNA are made up of introns and exons. According to their findings not all the nucleotides on the RNA strand (product of DNA transcription) are used in the translation process. The intervening sequences in the RNA strand are first spliced out so that only the RNA segment left behind after splicing would be translated to polypeptides.
- 1994: The first breast cancer gene is discovered. BRCA I, was discovered by researchers at the King laboratory at UC Berkeley in 1990 but was first cloned in 1994. BRCA II, the second key gene in the manifestation of breast cancer was discovered later in 1994 by Professor Michael Stratton and Dr. Richard Wooster.
- 1995: The genome of bacterium *Haemophilus influenzae* is the first genome of a free living organism to be sequenced
- 1996: *Saccharomyces cerevisiae*, a yeast species, is the first eukaryote genome sequence to be released
- 1996: Alexander Rich discovered the Z-DNA, a type of DNA which is in a transient state, that is in some cases associated with DNA transcription. The Z-DNA form is more likely to occur in regions of DNA rich in cytosine and guanine with high salt concentrations.
- 1997: Dolly the sheep was cloned by Ian Wilmut and colleagues from the Roslin Institute in Scotland.
- 1998: The first genome sequence for a multicellular eukaryote, *Caenorhabditis elegans*, is released
- 2000: The full genome sequence of *Drosophila melanogaster* is completed.
- 2001: First draft sequences of the human genome are released simultaneously by the Human Genome Project and Celera Genomics.

- 2001: Francisco Mojica and Rudd Jansen propose the acronym CRISPR to describe a family of bacterial DNA sequences that can be used to specifically change genes within organisms.
- 2003 (14 April): Successful completion of Human Genome Project with 99 per cent of the genome sequenced to a 99.99 per cent accuracy
- 2004: Merck introduced a vaccine for Human Papillomavirus which promised to protect women against infection with HPV 16 and 18, which inactivates tumor suppressor genes and together cause 70 per cent of cervical cancers.
- 2007: Michael Worobey traced the evolutionary origins of HIV by analyzing its genetic mutations, which revealed that HIV infections had occurred in the United States as early as the 1960s.
- 2007: Timothy Ray Brown becomes the first person cured from HIV/AIDS through a Hematopoietic stem cell transplantation.
- 2008: Houston-based Introgen developed Advexin (FDA Approval pending), the first gene therapy for cancer and Li-Fraumeni syndrome, utilizing a form of Adenovirus to carry a replacement gene coding for the p53 protein.
- 2010: transcription activator-like effector nucleases (or TALENs) are first used to cut specific sequences of DNA.
- 2016: A genome is sequenced in outer space for the first time, with NASA astronaut Kate Rubins using a MinION device aboard the International Space Station.

FEATURES OF INHERITANCE

Discrete inheritance and Mendel's laws

At its most fundamental level, inheritance in organisms occurs by passing discrete heritable units, called genes, from parents to offspring. This property was first observed by Gregor Mendel, who studied the segregation of heritable traits in pea plants. In his experiments studying the trait for flower colour, Mendel observed that the flowers of each pea plant were either purple or white—but never an intermediate between the two colors. These different, discrete versions of the same gene are called alleles. In the case of the pea, which is a diploid species, each individual plant has two copies of each gene, one copy inherited from each parent. Many species, including humans, have this pattern of inheritance. Diploid organisms with two copies of the same allele of a given gene are called homozygous at that gene locus, while organisms with two different alleles of a given gene are called heterozygous. The set of alleles for a given organism is called its genotype, while the observable traits of the organism are called its phenotype. When organisms are heterozygous at a gene, often one allele is called dominant as its qualities dominate the phenotype of the organism, while the other allele is called recessive as its qualities recede and are not observed. Some alleles do not have complete dominance and instead have incomplete dominance by expressing an intermediate phenotype, or codominance by expressing both alleles at once. When a pair of organisms reproduce sexually, their offspring randomly inherit one of the two alleles from each parent. These observations of discrete inheritance and the segregation of alleles are collectively known as Mendel's first law or the Law of Segregation.

NOTATION AND DIAGRAMS

Geneticists use diagrams and symbols to describe inheritance. A gene is represented by one or a few letters. Often a "+" symbol is used to mark the usual, non-mutant allele for a gene. In fertilization and breeding experiments (and especially when discussing Mendel's laws) the parents are referred to as the "P" generation and the offspring as the "F1" (first filial) generation. When the F1 offspring mate with each other, the offspring

are called the “F2” (second filial) generation. One of the common diagrams used to predict the result of cross-breeding is the Punnett square. When studying human genetic diseases, geneticists often use pedigree charts to represent the inheritance of traits. These charts map the inheritance of a trait in a family tree.

MULTIPLE GENE INTERACTIONS

Organisms have thousands of genes, and in sexually reproducing organisms these genes generally assort independently of each other. This means that the inheritance of an allele for yellow or green pea colour is unrelated to the inheritance of alleles for white or purple flowers. This phenomenon, known as “Mendel’s second law” or the “law of independent assortment,” means that the alleles of different genes get shuffled between parents to form offspring with many different combinations. (Some genes do not assort independently, demonstrating genetic linkage, a topic discussed later in this article.)

Often different genes can interact in a way that influences the same trait. In the Blue-eyed Mary (*Omphalodes verna*), for example, there exists a gene with alleles that determine the colour of flowers: blue or magenta. Another gene, however, controls whether the flowers have colour at all or are white. When a plant has two copies of this white allele, its flowers are white—regardless of whether the first gene has blue or magenta alleles. This interaction between genes is called epistasis, with the second gene epistatic to the first.

Many traits are not discrete features (*e.g.* purple or white flowers) but are instead continuous features (*e.g.* human height and skin colour). These complex traits are products of many genes. The influence of these genes is mediated, to varying degrees, by the environment an organism has experienced. The degree to which an organism’s genes contribute to a complex trait is called heritability. Measurement of the heritability of a trait is relative—in a more variable environment, the environment has a bigger influence on the total variation of the trait. For example, human height is a trait with complex causes. It has a heritability of 89 per cent in the United States. In Nigeria, however, where people experience a more variable access to good nutrition and health care, height has a heritability of only 62 per cent.

MOLECULAR BASIS FOR INHERITANCE

DNA and chromosomes

The molecular basis for genes is deoxyribonucleic acid (DNA). DNA is composed of a chain of nucleotides, of which there are four types: adenine (A), cytosine (C), guanine (G), and thymine (T). Genetic information exists in the sequence of these nucleotides, and genes exist as stretches of sequence along the DNA chain. Viruses are the only exception to this rule—sometimes viruses use the very similar molecule RNA instead of DNA as their genetic material. Viruses cannot reproduce without a host and are unaffected by many genetic processes, so tend not to be considered living organisms. DNA normally exists as a double-stranded molecule, coiled into the shape of a double helix. Each nucleotide in DNA preferentially pairs with its partner nucleotide on the opposite strand: A pairs with T, and C pairs with G. Thus, in its two-stranded form, each strand effectively contains all necessary information, redundant with its partner strand. This structure of DNA is the physical basis for inheritance: DNA replication duplicates the genetic information by splitting the strands and using each strand as a template for synthesis of a new partner strand.

Genes are arranged linearly along long chains of DNA base-pair sequences. In bacteria, each cell usually contains a single circular genophore, while eukaryotic organisms (such as plants and animals) have their DNA arranged in multiple linear chromosomes. These DNA strands are often extremely long; the largest human chromosome, for example, is about 247 million base pairs in length. The DNA of a chromosome is associated with structural proteins that organize, compact, and control access to the DNA, forming a material called

chromatin; in eukaryotes, chromatin is usually composed of nucleosomes, segments of DNA wound around cores of histone proteins. The full set of hereditary material in an organism (usually the combined DNA sequences of all chromosomes) is called the genome. While haploid organisms have only one copy of each chromosome, most animals and many plants are diploid, containing two of each chromosome and thus two copies of every gene. The two alleles for a gene are located on identical loci of the two homologous chromosomes, each allele inherited from a different parent. Many species have so-called sex chromosomes that determine the gender of each organism. In humans and many other animals, the Y chromosome contains the gene that triggers the development of the specifically male characteristics. In evolution, this chromosome has lost most of its content and also most of its genes, while the X chromosome is similar to the other chromosomes and contains many genes. The X and Y chromosomes form a strongly heterogeneous pair.

REPRODUCTION

When cells divide, their full genome is copied and each daughter cell inherits one copy. This process, called mitosis, is the simplest form of reproduction and is the basis for asexual reproduction. Asexual reproduction can also occur in multicellular organisms, producing offspring that inherit their genome from a single parent. Offspring that are genetically identical to their parents are called clones. Eukaryotic organisms often use sexual reproduction to generate offspring that contain a mixture of genetic material inherited from two different parents. The process of sexual reproduction alternates between forms that contain single copies of the genome (haploid) and double copies (diploid). Haploid cells fuse and combine genetic material to create a diploid cell with paired chromosomes. Diploid organisms form haploids by dividing, without replicating their DNA, to create daughter cells that randomly inherit one of each pair of chromosomes. Most animals and many plants are diploid for most of their lifespan, with the haploid form reduced to single cell gametes such as sperm or eggs. Although they do not use the haploid/diploid method of sexual reproduction, bacteria have many methods of acquiring new genetic information. Some bacteria can undergo conjugation, transferring a small circular piece of DNA to another bacterium. Bacteria can also take up raw DNA fragments found in the environment and integrate them into their genomes, a phenomenon known as transformation. These processes result in horizontal gene transfer, transmitting fragments of genetic information between organisms that would be otherwise unrelated.

RECOMBINATION AND GENETIC LINKAGE

The diploid nature of chromosomes allows for genes on different chromosomes to assort independently or be separated from their homologous pair during sexual reproduction wherein haploid gametes are formed. In this way new combinations of genes can occur in the offspring of a mating pair. Genes on the same chromosome would theoretically never recombine. However, they do, via the cellular process of chromosomal crossover. During crossover, chromosomes exchange stretches of DNA, effectively shuffling the gene alleles between the chromosomes. This process of chromosomal crossover generally occurs during meiosis, a series of cell divisions that creates haploid cells. The first cytological demonstration of crossing over was performed by Harriet Creighton and Barbara McClintock in 1931. Their research and experiments on corn provided cytological evidence for the genetic theory that linked genes on paired chromosomes do in fact exchange places from one homolog to the other.

The probability of chromosomal crossover occurring between two given points on the chromosome is related to the distance between the points. For an arbitrarily long distance, the probability of crossover is high enough that the inheritance of the genes is effectively uncorrelated. For genes that are closer together, however, the lower probability of crossover means that the genes demonstrate genetic linkage; alleles for the two genes tend to be inherited together. The amounts of linkage between a series of genes can be combined to form a linear linkage map that roughly describes the arrangement of the genes along the chromosome.

GENE EXPRESSION

Genetic code

Genes generally express their functional effect through the production of proteins, which are complex molecules responsible for most functions in the cell. Proteins are made up of one or more polypeptide chains, each of which is composed of a sequence of amino acids, and the DNA sequence of a gene (through an RNA intermediate) is used to produce a specific amino acid sequence. This process begins with the production of an RNA molecule with a sequence matching the gene's DNA sequence, a process called transcription. This messenger RNA molecule is then used to produce a corresponding amino acid sequence through a process called translation. Each group of three nucleotides in the sequence, called a codon, corresponds either to one of the twenty possible amino acids in a protein or an instruction to end the amino acid sequence; this correspondence is called the genetic code. The flow of information is unidirectional: information is transferred from nucleotide sequences into the amino acid sequence of proteins, but it never transfers from protein back into the sequence of DNA—a phenomenon Francis Crick called the central dogma of molecular biology. The specific sequence of amino acids results in a unique three-dimensional structure for that protein, and the three-dimensional structures of proteins are related to their functions. Some are simple structural molecules, like the fibers formed by the protein collagen. Proteins can bind to other proteins and simple molecules, sometimes acting as enzymes by facilitating chemical reactions within the bound molecules (without changing the structure of the protein itself). Protein structure is dynamic; the protein hemoglobin bends into slightly different forms as it facilitates the capture, transport, and release of oxygen molecules within mammalian blood.

A single nucleotide difference within DNA can cause a change in the amino acid sequence of a protein. Because protein structures are the result of their amino acid sequences, some changes can dramatically change the properties of a protein by destabilizing the structure or changing the surface of the protein in a way that changes its interaction with other proteins and molecules. For example, sickle-cell anemia is a human genetic disease that results from a single base difference within the coding region for the β -globin section of hemoglobin, causing a single amino acid change that changes hemoglobin's physical properties. Sickle-cell versions of hemoglobin stick to themselves, stacking to form fibers that distort the shape of red blood cells carrying the protein. These sickle-shaped cells no longer flow smoothly through blood vessels, having a tendency to clog or degrade, causing the medical problems associated with this disease. Some DNA sequences are transcribed into RNA but are not translated into protein products—such RNA molecules are called non-coding RNA. In some cases, these products fold into structures which are involved in critical cell functions (*e.g.* ribosomal RNA and transfer RNA). RNA can also have regulatory effects through hybridization interactions with other RNA molecules (*e.g.* microRNA).

NATURE AND NURTURE

Although genes contain all the information an organism uses to function, the environment plays an important role in determining the ultimate phenotypes an organism displays. The phrase “nature and nurture” refers to this complementary relationship. The phenotype of an organism depends on the interaction of genes and the environment. An interesting example is the coat coloration of the Siamese cat. In this case, the body temperature of the cat plays the role of the environment. The cat's genes code for dark hair, thus the hair-producing cells in the cat make cellular proteins resulting in dark hair. But these dark hair-producing proteins are sensitive to temperature (*i.e.* have a mutation causing temperature-sensitivity) and denature in higher-temperature environments, failing to produce dark-hair pigment in areas where the cat has a higher body temperature. In a low-temperature environment, however, the protein's structure is stable and produces dark-hair pigment normally.

The protein remains functional in areas of skin that are colder—such as its legs, ears, tail and face—so the cat has dark-hair at its extremities. Environment plays a major role in effects of the human genetic disease phenylketonuria. The mutation that causes phenylketonuria disrupts the ability of the body to break down the amino acid phenylalanine, causing a toxic build-up of an intermediate molecule that, in turn, causes severe symptoms of progressive intellectual disability and seizures. However, if someone with the phenylketonuria mutation follows a strict diet that avoids this amino acid, they remain normal and healthy. A common method for determining how genes and environment (“nature and nurture”) contribute to a phenotype involves studying identical and fraternal twins, or other siblings of multiple births. Because identical siblings come from the same zygote, they are genetically the same. Fraternal twins are as genetically different from one another as normal siblings. By comparing how often a certain disorder occurs in a pair of identical twins to how often it occurs in a pair of fraternal twins, scientists can determine whether that disorder is caused by genetic or postnatal environmental factors – whether it has “nature” or “nurture” causes. One famous example involved the study of the Genain quadruplets, who were identical quadruplets all diagnosed with schizophrenia. However such tests cannot separate genetic factors from environmental factors affecting fetal development.

GENE REGULATION

The genome of a given organism contains thousands of genes, but not all these genes need to be active at any given moment. A gene is expressed when it is being transcribed into mRNA and there exist many cellular methods of controlling the expression of genes such that proteins are produced only when needed by the cell. Transcription factors are regulatory proteins that bind to DNA, either promoting or inhibiting the transcription of a gene. Within the genome of *Escherichia coli* bacteria, for example, there exists a series of genes necessary for the synthesis of the amino acid tryptophan. However, when tryptophan is already available to the cell, these genes for tryptophan synthesis are no longer needed. The presence of tryptophan directly affects the activity of the genes—tryptophan molecules bind to the tryptophan repressor (a transcription factor), changing the repressor’s structure such that the repressor binds to the genes. The tryptophan repressor blocks the transcription and expression of the genes, thereby creating negative feedback regulation of the tryptophan synthesis process. Differences in gene expression are especially clear within multicellular organisms, where cells all contain the same genome but have very different structures and behaviors due to the expression of different sets of genes. All the cells in a multicellular organism derive from a single cell, differentiating into variant cell types in response to external and intercellular signals and gradually establishing different patterns of gene expression to create different behaviors. As no single gene is responsible for the development of structures within multicellular organisms, these patterns arise from the complex interactions between many cells. Within eukaryotes, there exist structural features of chromatin that influence the transcription of genes, often in the form of modifications to DNA and chromatin that are stably inherited by daughter cells. These features are called “epigenetic” because they exist “on top” of the DNA sequence and retain inheritance from one cell generation to the next. Because of epigenetic features, different cell types grown within the same medium can retain very different properties. Although epigenetic features are generally dynamic over the course of development, some, like the phenomenon of paramutation, have multigenerational inheritance and exist as rare exceptions to the general rule of DNA as the basis for inheritance.

GENETIC CHANGE

Mutations

In biology, a mutation is the permanent alteration of the nucleotide sequence of the genome of an organism, virus, or extrachromosomal DNA or other genetic elements.

Mutations result from errors during DNA replication (especially during meiosis) or other types of damage to DNA (such as may be caused by exposure to radiation or carcinogens), which then may undergo error-prone repair (especially [microhomology-mediated end joining]), or cause an error during other forms of repair, or else may cause an error during replication (translesion synthesis). Mutations may also result from insertion or deletion of segments of DNA due to mobile genetic elements. Mutations may or may not produce discernible changes in the observable characteristics (phenotype) of an organism. Mutations play a part in both normal and abnormal biological processes including: evolution, cancer, and the development of the immune system, including junctional diversity. The genomes of RNA viruses are based on RNA rather than DNA. The RNA viral genome can be double stranded (as in DNA) or single stranded. In some of these viruses (such as the single stranded human immunodeficiency virus) replication occurs quickly and there are no mechanisms to check the genome for accuracy. This error-prone process often results in mutations. Mutation can result in many different types of change in sequences. Mutations in genes can either have no effect, alter the product of a gene, or prevent the gene from functioning properly or completely. Mutations can also occur in nongenic regions. One study on genetic variations between different species of *Drosophila* suggests that, if a mutation changes a protein produced by a gene, the result is likely to be harmful, with an estimated 70 percent of amino acid polymorphisms that have damaging effects, and the remainder being either neutral or marginally beneficial. Due to the damaging effects that mutations can have on genes, organisms have mechanisms such as DNA repair to prevent or correct mutations by reverting the mutated sequence back to its original state.

Description

Mutations can involve the duplication of large sections of DNA, usually through genetic recombination. These duplications are a major source of raw material for evolving new genes, with tens to hundreds of genes duplicated in animal genomes every million years. Most genes belong to larger gene families of shared ancestry, known as homology. Novel genes are produced by several methods, commonly through the duplication and mutation of an ancestral gene, or by recombining parts of different genes to form new combinations with new functions. Here, protein domains act as modules, each with a particular and independent function, that can be mixed together to produce genes encoding new proteins with novel properties. For example, the human eye uses four genes to make structures that sense light: three for cone cell or colour vision and one for rod cell or night vision; all four arose from a single ancestral gene. Another advantage of duplicating a gene (or even an entire genome) is that this increases engineering redundancy; this allows one gene in the pair to acquire a new function while the other copy performs the original function. Other types of mutation occasionally create new genes from previously noncoding DNA. Changes in chromosome number may involve even larger mutations, where segments of the DNA within chromosomes break and then rearrange. For example, in the Hominae, two chromosomes fused to produce human chromosome 2; this fusion did not occur in the lineage of the other apes, and they retain these separate chromosomes. In evolution, the most important role of such chromosomal rearrangements may be to accelerate the divergence of a population into new species by making populations less likely to interbreed, thereby preserving genetic differences between these populations.

Sequences of DNA that can move about the genome, such as transposons, make up a major fraction of the genetic material of plants and animals, and may have been important in the evolution of genomes. For example, more than a million copies of the Alu sequence are present in the human genome, and these sequences have now been recruited to perform functions such as regulating gene expression. Another effect of these mobile DNA sequences is that when they move within a genome, they can mutate or delete existing genes and thereby produce genetic diversity. Nonlethal mutations accumulate within the gene pool and increase the amount of genetic variation. The abundance of some genetic changes within the gene pool can be reduced by natural selection, while other “more favorable” mutations may accumulate and result in adaptive changes.

For example, a butterfly may produce offspring with new mutations. The majority of these mutations will have no effect; but one might change the colour of one of the butterfly's offspring, making it harder (or easier) for predators to see. If this colour change is advantageous, the chances of this butterfly's surviving and producing its own offspring are a little better, and over time the number of butterflies with this mutation may form a larger percentage of the population. Neutral mutations are defined as mutations whose effects do not influence the fitness of an individual. These can increase in frequency over time due to genetic drift. It is believed that the overwhelming majority of mutations have no significant effect on an organism's fitness. Also, DNA repair mechanisms are able to mend most changes before they become permanent mutations, and many organisms have mechanisms for eliminating otherwise-permanently mutated somatic cells.

Beneficial mutations can improve reproductive success.

History

Mutationism is one of several alternatives to evolution by natural selection that have existed both before and after the publication of Charles Darwin's 1859 book, *On the Origin of Species*. In the theory, mutation was the source of novelty, creating new forms and new species, potentially instantaneously, in a sudden jump. This was envisaged as driving evolution, which was limited by the supply of mutations. Before Darwin, biologists commonly believed in saltationism, the possibility of large evolutionary jumps, including immediate speciation. For example, in 1822 Étienne Geoffroy Saint-Hilaire argued that species could be formed by sudden transformations, or what would later be called macromutation. Darwin opposed saltation, insisting on gradualism in evolution as in geology. In 1864, Albert von K  lliker revived Geoffroy's theory. In 1901 the geneticist Hugo de Vries gave the name "mutation" to seemingly new forms that suddenly arose in his experiments on the evening primrose *Oenothera lamarckiana*, and in the first decade of the 20th century, mutationism, or as de Vries named it *mutationstheorie*, became a rival to Darwinism supported for a while by geneticists including William Bateson, Thomas Hunt Morgan, and Reginald Punnett. Understanding of mutationism is clouded by the mid-20th century portrayal of the early mutationists by supporters of the modern synthesis as opponents of Darwinian evolution and rivals of the biometrics school who argued that selection operated on continuous variation. In this portrayal, mutationism was defeated by a synthesis of genetics and natural selection that supposedly started later, around 1918, with work by the mathematician Ronald Fisher. However, the alignment of Mendelian genetics and natural selection began as early as 1902 with a paper by Udny Yule, and built up with theoretical and experimental work in Europe and America. Despite the controversy, the early mutationists had by 1918 already accepted natural selection and explained continuous variation as the result of multiple genes acting on the same characteristic, such as height. Mutationism, along with other alternatives to Darwinism like Lamarckism and orthogenesis, was discarded by most biologists as they came to see that Mendelian genetics and natural selection could readily work together; mutation took its place as a source of the genetic variation essential for natural selection to work on. However, mutationism did not entirely vanish. In 1940, Richard Goldschmidt again argued for single-step speciation by macromutation, describing the organisms thus produced as "hopeful monsters", earning widespread ridicule. In 1987, Masatoshi Nei argued controversially that evolution was often mutation-limited. Modern biologists such as Douglas J. Futuyma conclude that essentially all claims of evolution driven by large mutations can be explained by Darwinian evolution.

Causes

Four classes of mutations are:

- Spontaneous mutations (molecular decay),
- Mutations due to error-prone replication bypass of naturally occurring DNA damage (also called error-prone translesion synthesis),

- Errors introduced during DNA repair, and
- Induced mutations caused by mutagens. Scientists may also deliberately introduce mutant sequences through DNA manipulation for the sake of scientific experimentation.

One 2017 study claimed that 66 per cent of cancer-causing mutations are random, 29 per cent are due to the environment (the studied population spanned 69 countries), and 5 per cent are inherited.

Humans on average pass 60 new mutations to their children but fathers pass more mutations depending on their age with every year adding two new mutations to a child.

Spontaneous Mutation

Spontaneous mutations occur with non-zero probability even given a healthy, uncontaminated cell. They can be characterized by the specific change:

- Tautomerism — A base is changed by the repositioning of a hydrogen atom, altering the hydrogen bonding pattern of that base, resulting in incorrect base pairing during replication.
- Depurination — Loss of a purine base (A or G) to form an apurinic site (AP site).
- Deamination — Hydrolysis changes a normal base to an atypical base containing a keto group in place of the original amine group. Examples include C → U and A → HX (hypoxanthine), which can be corrected by DNA repair mechanisms; and 5MeC (5-methylcytosine) → T, which is less likely to be detected as a mutation because thymine is a normal DNA base.
- Slipped strand mispairing — Denaturation of the new strand from the template during replication, followed by renaturation in a different spot (“slipping”). This can lead to insertions or deletions.
- Replication slippage

Error-prone Replication Bypass

There is increasing evidence that the majority of spontaneously arising mutations are due to error-prone replication (translesion synthesis) past DNA damage in the template strand. Naturally occurring oxidative DNA damages arise at least 10,000 times per cell per day in humans and 50,000 times or more per cell per day in rats.

In mice, the majority of mutations are caused by translesion synthesis. Likewise, in yeast, Kunz et al. found that more than 60 per cent of the spontaneous single base pair substitutions and deletions were caused by translation synthesis.

Errors Introduced During DNA Repair

Although naturally occurring double-strand breaks occur at a relatively low frequency in DNA, their repair often causes mutation. Non-homologous end joining (NHEJ) is a major pathway for repairing double-strand breaks. NHEJ involves removal of a few nucleotides to allow somewhat inaccurate alignment of the two ends for rejoining followed by addition of nucleotides to fill in gaps. As a consequence, NHEJ often introduces mutations.

Induced Mutation

Induced mutations are alterations in the gene after it has come in contact with mutagens and environmental causes.

Induced mutations on the molecular level can be caused by:

- Chemicals
 - (a) Hydroxylamine

- (b) Base analogs (*e.g.*, Bromodeoxyuridine (BrdU))
- (c) Alkylating agents (*e.g.*, *N*-ethyl-*N*-nitrosourea (ENU)). These agents can mutate both replicating and non-replicating DNA. In contrast, a base analog can mutate the DNA only when the analog is incorporated in replicating the DNA. Each of these classes of chemical mutagens has certain effects that then lead to transitions, transversions, or deletions.
- (d) Agents that form DNA adducts (*e.g.*, ochratoxin A)
- (e) DNA intercalating agents (*e.g.*, ethidium bromide)
- DNA crosslinkers
 - (a) Oxidative damage
 - (b) Nitrous acid converts amine groups on A and C to diazo groups, altering their hydrogen bonding patterns, which leads to incorrect base pairing during replication.
- Radiation
 - (a) Ultraviolet light (UV) (non-ionizing radiation). Two nucleotide bases in DNA—cytosine and thymine—are most vulnerable to radiation that can change their properties. UV light can induce adjacent pyrimidine bases in a DNA strand to become covalently joined as a pyrimidine dimer. UV radiation, in particular longer-wave UVA, can also cause oxidative damage to DNA.
 - (b) Ionizing radiation. Exposure to ionizing radiation, such as gamma radiation, can result in mutation, possibly resulting in cancer or death.

Classification of Types

By Effect on Structure

The sequence of a gene can be altered in a number of ways. Gene mutations have varying effects on health depending on where they occur and whether they alter the function of essential proteins. Mutations in the structure of genes can be classified into several types.

Small-scale Mutations

Small-scale mutations affect a gene in one or a few nucleotides. (If only a single nucleotide is affected, they are called point mutations.) Small-scale mutations include:

- Insertions add one or more extra nucleotides into the DNA. They are usually caused by transposable elements, or errors during replication of repeating elements. Insertions in the coding region of a gene may alter splicing of the mRNA (splice site mutation), or cause a shift in the reading frame (frameshift), both of which can significantly alter the gene product. Insertions can be reversed by excision of the transposable element.
- Deletions remove one or more nucleotides from the DNA. Like insertions, these mutations can alter the reading frame of the gene. In general, they are irreversible: Though exactly the same sequence might in theory be restored by an insertion, transposable elements able to revert a very short deletion (say 1–2 bases) in *any* location either are highly unlikely to exist or do not exist at all.
- Substitution mutations, often caused by chemicals or malfunction of DNA replication, exchange a single nucleotide for another. These changes are classified as transitions or transversions. Most common is the transition that exchanges a purine for a purine ($A \leftrightarrow G$) or a pyrimidine for a pyrimidine, ($C \leftrightarrow T$). A transition can be caused by nitrous acid, base mis-pairing, or mutagenic base analogs such as BrdU. Less common is a transversion, which exchanges a purine for a pyrimidine or a pyrimidine for a purine ($C/T \leftrightarrow A/G$). An example of a transversion is the conversion of adenine (A) into a cytosine (C). A point mutation are modifications of single base pairs of DNA or other small base pairs within a

gene. A point mutation can be reversed by another point mutation, in which the nucleotide is changed back to its original state (true reversion) or by second-site reversion (a complementary mutation elsewhere that results in regained gene functionality). As discussed below, point mutations that occur within the protein coding region of a gene may be classified as synonymous or nonsynonymous substitutions, the latter of which in turn can be divided into missense or nonsense mutations.

Large-scale Mutations

Large-scale mutations in chromosomal structure include:

- Amplifications (or gene duplications) leading to multiple copies of all chromosomal regions, increasing the dosage of the genes located within them.
- Deletions of large chromosomal regions, leading to loss of the genes within those regions.
- Mutations whose effect is to juxtapose previously separate pieces of DNA, potentially bringing together separate genes to form functionally distinct fusion genes (*e.g.*, bcr-abl).
- Large scale changes to the structure of chromosomes called chromosomal rearrangement that can lead to a decrease of fitness but also to speciation in isolated, inbred populations. These include:
 - (a) Chromosomal translocations: interchange of genetic parts from nonhomologous chromosomes.
 - (b) Chromosomal inversions: reversing the orientation of a chromosomal segment.
 - (c) Non-homologous chromosomal crossover.
 - (d) Interstitial deletions: an intra-chromosomal deletion that removes a segment of DNA from a single chromosome, thereby apposing previously distant genes. For example, cells isolated from a human astrocytoma, a type of brain tumor, were found to have a chromosomal deletion removing sequences between the Fused in Glioblastoma (FIG) gene and the receptor tyrosine kinase (ROS), producing a fusion protein (FIG-ROS). The abnormal FIG-ROS fusion protein has constitutively active kinase activity that causes oncogenic transformation (a transformation from normal cells to cancer cells).
- Loss of heterozygosity: loss of one allele, either by a deletion or a genetic recombination event, in an organism that previously had two different alleles.

By Effect on Function

- Loss-of-function mutations, also called inactivating mutations, result in the gene product having less or no function (being partially or wholly inactivated). When the allele has a complete loss of function (null allele), it is often called an amorph or amorphic mutation in the Muller's morphs schema. Phenotypes associated with such mutations are most often recessive. Exceptions are when the organism is haploid, or when the reduced dosage of a normal gene product is not enough for a normal phenotype (this is called haploinsufficiency).
- Gain-of-function mutations, also called activating mutations, change the gene product such that its effect gets stronger (enhanced activation) or even is superseded by a different and abnormal function. When the new allele is created, a heterozygote containing the newly created allele as well as the original will express the new allele; genetically this defines the mutations as dominant phenotypes. Several of Muller's morphs correspond to gain of function, including hypermorph and neomorph. In December 2017, the U.S. government lifted a temporary ban implemented in 2014 that banned federal funding for any new "gain-of-function" experiments that enhance pathogens "such as Avian influenza, SARS and the Middle East Respiratory Syndrome or MERS viruses."
- Dominant negative mutations (also called antimorphic mutations) have an altered gene product that acts antagonistically to the wild-type allele. These mutations usually result in an altered molecular

function (often inactive) and are characterized by a dominant or semi-dominant phenotype. In humans, dominant negative mutations have been implicated in cancer (*e.g.*, mutations in genes p53, ATM, CEBPA and PPARGgamma). Marfan syndrome is caused by mutations in the *FBN1* gene, located on chromosome 15, which encodes fibrillin-1, a glycoprotein component of the extracellular matrix. Marfan syndrome is also an example of dominant negative mutation and haploinsufficiency.

- Hypomorphs, after Mullerian classification, are characterized by altered gene products that acts with decreased gene expression compared to the wild type allele.
- Neomorphs are characterized by the control of new protein product synthesis.
- Lethal mutations are mutations that lead to the death of the organisms that carry the mutations.
- A back mutation or reversion is a point mutation that restores the original sequence and hence the original phenotype.

By Effect on Fitness

In applied genetics, it is usual to speak of mutations as either harmful or beneficial.

- A harmful, or deleterious, mutation decreases the fitness of the organism.
- A beneficial, or advantageous mutation increases the fitness of the organism. Mutations that promotes traits that are desirable, are also called beneficial. In theoretical population genetics, it is more usual to speak of mutations as deleterious or advantageous than harmful or beneficial.
- A neutral mutation has no harmful or beneficial effect on the organism. Such mutations occur at a steady rate, forming the basis for the molecular clock. In the neutral theory of molecular evolution, neutral mutations provide genetic drift as the basis for most variation at the molecular level.
- A nearly neutral mutation is a mutation that may be slightly deleterious or advantageous, although most nearly neutral mutations are slightly deleterious.

Distribution of Fitness Effects

Attempts have been made to infer the distribution of fitness effects (DFE) using mutagenesis experiments and theoretical models applied to molecular sequence data. DFE, as used to determine the relative abundance of different types of mutations (*i.e.*, strongly deleterious, nearly neutral or advantageous), is relevant to many evolutionary questions, such as the maintenance of genetic variation, the rate of genomic decay, the maintenance of outcrossing sexual reproduction as opposed to inbreeding and the evolution of sex and genetic recombination. In summary, the DFE plays an important role in predicting evolutionary dynamics. A variety of approaches have been used to study the DFE, including theoretical, experimental and analytical methods.

- *Mutagenesis experiment:* The direct method to investigate the DFE is to induce mutations and then measure the mutational fitness effects, which has already been done in viruses, bacteria, yeast, and *Drosophila*. For example, most studies of the DFE in viruses used site-directed mutagenesis to create point mutations and measure relative fitness of each mutant. In *Escherichia coli*, one study used transposon mutagenesis to directly measure the fitness of a random insertion of a derivative of Tn10. In yeast, a combined mutagenesis and deep sequencing approach has been developed to generate high-quality systematic mutant libraries and measure fitness in high throughput. However, given that many mutations have effects too small to be detected and that mutagenesis experiments can detect only mutations of moderately large effect; DNA sequence data analysis can provide valuable information about these mutations.
- *Molecular sequence analysis:* With rapid development of DNA sequencing technology, an enormous amount of DNA sequence data is available and even more is forthcoming in the future. Various methods have been developed to infer the DFE from DNA sequence data. By examining DNA

sequence differences within and between species, we are able to infer various characteristics of the DFE for neutral, deleterious and advantageous mutations. To be specific, the DNA sequence analysis approach allows us to estimate the effects of mutations with very small effects, which are hardly detectable through mutagenesis experiments.

One of the earliest theoretical studies of the distribution of fitness effects was done by Motoo Kimura, an influential theoretical population geneticist. His neutral theory of molecular evolution proposes that most novel mutations will be highly deleterious, with a small fraction being neutral. Hiroshi Akashi more recently proposed a bimodal model for the DFE, with modes centered around highly deleterious and neutral mutations. Both theories agree that the vast majority of novel mutations are neutral or deleterious and that advantageous mutations are rare, which has been supported by experimental results. One example is a study done on the DFE of random mutations in vesicular stomatitis virus. Out of all mutations, 39.6 per cent were lethal, 31.2 per cent were non-lethal deleterious, and 27.1 per cent were neutral. Another example comes from a high throughput mutagenesis experiment with yeast. In this experiment it was shown that the overall DFE is bimodal, with a cluster of neutral mutations, and a broad distribution of deleterious mutations. Though relatively few mutations are advantageous, those that are play an important role in evolutionary changes. Like neutral mutations, weakly selected advantageous mutations can be lost due to random genetic drift, but strongly selected advantageous mutations are more likely to be fixed. Knowing the DFE of advantageous mutations may lead to increased ability to predict the evolutionary dynamics. Theoretical work on the DFE for advantageous mutations has been done by John H. Gillespie and H. Allen Orr. They proposed that the distribution for advantageous mutations should be exponential under a wide range of conditions, which, in general, has been supported by experimental studies, at least for strongly selected advantageous mutations. In general, it is accepted that the majority of mutations are neutral or deleterious, with advantageous mutations being rare; however, the proportion of types of mutations varies between species. This indicates two important points: first, the proportion of effectively neutral mutations is likely to vary between species, resulting from dependence on effective population size; second, the average effect of deleterious mutations varies dramatically between species. In addition, the DFE also differs between coding regions and noncoding regions, with the DFE of noncoding DNA containing more weakly selected mutations.

By Impact on Protein Sequence

- A frameshift mutation is a mutation caused by insertion or deletion of a number of nucleotides that is not evenly divisible by three from a DNA sequence. Due to the triplet nature of gene expression by codons, the insertion or deletion can disrupt the reading frame, or the grouping of the codons, resulting in a completely different translation from the original. The earlier in the sequence the deletion or insertion occurs, the more altered the protein produced is. (For example, the code CCU GAC UAC CUA codes for the amino acids proline, aspartic acid, tyrosine, and leucine. If the U in CCU was deleted, the resulting sequence would be CCG ACU ACC UAx, which would instead code for proline, threonine, threonine, and part of another amino acid or perhaps a stop codon (where the x stands for the following nucleotide).) By contrast, any insertion or deletion that is evenly divisible by three is termed an *in-frame mutation*.
- A point substitution mutation results in a change in a single nucleotide and can be either synonymous or nonsynonymous.
 - (a) A synonymous substitution replaces a codon with another codon that codes for the same amino acid, so that the produced amino acid sequence is not modified. Synonymous mutations occur due to the degenerate nature of the genetic code. If this mutation does not result in any phenotypic effects, then it is called silent, but not all synonymous substitutions are silent. (There can also

be silent mutations in nucleotides outside of the coding regions, such as the introns, because the exact nucleotide sequence is not as crucial as it is in the coding regions, but these are not considered synonymous substitutions.)

- (b) A nonsynonymous substitution replaces a codon with another codon that codes for a different amino acid, so that the produced amino acid sequence is modified. Nonsynonymous substitutions can in turn be classified as nonsense or missense mutations:
- A missense mutation or changes a nucleotide is to cause substitution of a different amino acid. This in turn can render the resulting protein nonfunctional. Such mutations are responsible for diseases such as Epidermolysis bullosa, sickle-cell disease, and SOD1-mediated ALS. On the other hand, if a missense mutation occurs in an amino acid codon that results in the use of a different, but chemically similar, amino acid, then sometimes little or no change is rendered in the protein. For example, a change from AAA to AGA will encode arginine, a chemically similar molecule to the intended lysine. In this latter case the mutation will have little or no effect on phenotype and therefore be neutral.
 - A nonsense mutation is a point mutation in a sequence of DNA that results in a premature stop codon, or a *nonsense codon* in the transcribed mRNA, and possibly a truncated, and often nonfunctional protein product. This sort of mutation has been linked to different mutations, such as congenital adrenal hyperplasia.

By Inheritance

In multicellular organisms with dedicated reproductive cells, mutations can be subdivided into germline mutations, which can be passed on to descendants through their reproductive cells, and somatic mutations (also called acquired mutations), which involve cells outside the dedicated reproductive group and which are not usually transmitted to descendants. A germline mutation gives rise to a *constitutional mutation* in the offspring, that is, a mutation that is present in every cell. A constitutional mutation can also occur very soon after fertilisation, or continue from a previous constitutional mutation in a parent. The distinction between germline and somatic mutations is important in animals that have a dedicated germline to produce reproductive cells. However, it is of little value in understanding the effects of mutations in plants, which lack dedicated germline. The distinction is also blurred in those animals that reproduce asexually through mechanisms such as budding, because the cells that give rise to the daughter organisms also give rise to that organism's germline. A new germline mutation not inherited from either parent is called a *de novo* mutation.

Diploid organisms (*e.g.*, humans) contain two copies of each gene—a paternal and a maternal allele. Based on the occurrence of mutation on each chromosome, we may classify mutations into three types.

- A heterozygous mutation is a mutation of only one allele.
- A homozygous mutation is an identical mutation of both the paternal and maternal alleles.
- Compound heterozygous mutations or a genetic compound consists of two different mutations in the paternal and maternal alleles.

A wild type or homozygous non-mutated organism is one in which neither allele is mutated.

Special Classes

- Conditional mutation is a mutation that has wild-type (or less severe) phenotype under certain “permissive” environmental conditions and a mutant phenotype under certain “restrictive” conditions. For example, a temperature-sensitive mutation can cause cell death at high temperature (restrictive condition), but might have no deleterious consequences at a lower temperature (permissive condition). These mutations are non-autonomous, as their manifestation depends upon presence of certain

conditions, as opposed to other mutations which appear autonomously. The permissive conditions may be temperature, certain chemicals, light or mutations in other parts of the genome. *In vivo* mechanisms like transcriptional switches can create conditional mutations. For instance, association of Steroid Binding Domain can create a transcriptional switch that can change the expression of a gene based on the presence of a steroid ligand. Conditional mutations have applications in research as they allow control over gene expression. This is especially useful studying diseases in adults by allowing expression after a certain period of growth, thus eliminating the deleterious effect of gene expression seen during stages of development in model organisms. DNA Recombinase systems like Cre-Lox Recombination used in association with promoters that are activated under certain conditions can generate conditional mutations. Dual Recombinase technology can be used to induce multiple conditional mutations to study the diseases which manifest as a result of simultaneous mutations in multiple genes. Certain inteins have been identified which splice only at certain permissive temperatures, leading to improper protein synthesis and thus, loss of function mutations at other temperatures. Conditional mutations may also be used in genetic studies associated with ageing, as the expression can be changed after a certain time period in the organism's lifespan.

- Replication timing quantitative trait loci affects DNA replication.

Nomenclature

In order to categorize a mutation as such, the “normal” sequence must be obtained from the DNA of a “normal” or “healthy” organism (as opposed to a “mutant” or “sick” one), it should be identified and reported; ideally, it should be made publicly available for a straightforward nucleotide-by-nucleotide comparison, and agreed upon by the scientific community or by a group of expert geneticists and biologists, who have the responsibility of establishing the *standard* or so-called “consensus” sequence. This step requires a tremendous scientific effort. Once the consensus sequence is known, the mutations in a genome can be pinpointed, described, and classified. The committee of the Human Genome Variation Society (HGVS) has developed the standard human sequence variant nomenclature, which should be used by researchers and DNA diagnostic centers to generate unambiguous mutation descriptions. In principle, this nomenclature can also be used to describe mutations in other organisms. The nomenclature specifies the type of mutation and base or amino acid changes.

- Nucleotide substitution (*e.g.*, 76A>T) — The number is the position of the nucleotide from the 5' end; the first letter represents the wild-type nucleotide, and the second letter represents the nucleotide that replaced the wild type. In the given example, the adenine at the 76th position was replaced by a thymine.
 - (a) If it becomes necessary to differentiate between mutations in genomic DNA, mitochondrial DNA, and RNA, a simple convention is used. For example, if the 100th base of a nucleotide sequence mutated from G to C, then it would be written as g.100G>C if the mutation occurred in genomic DNA, m.100G>C if the mutation occurred in mitochondrial DNA, or r.100g>c if the mutation occurred in RNA. Note that, for mutations in RNA, the nucleotide code is written in lower case.
- Amino acid substitution (*e.g.*, D111E) — The first letter is the one letter code of the wild-type amino acid, the number is the position of the amino acid from the N-terminus, and the second letter is the one letter code of the amino acid present in the mutation. Nonsense mutations are represented with an X for the second amino acid (*e.g.* D111X).
- Amino acid deletion (*e.g.*, ΔF508) — The Greek letter Δ (delta) indicates a deletion. The letter refers to the amino acid present in the wild type and the number is the position from the N terminus of the amino acid were it to be present as in the wild type.

Mutation Rates

Mutation rates vary substantially across species, and the evolutionary forces that generally determine mutation are the subject of ongoing investigation.

Harmful Mutations

Changes in DNA caused by mutation can cause errors in protein sequence, creating partially or completely non-functional proteins. Each cell, in order to function correctly, depends on thousands of proteins to function in the right places at the right times. When a mutation alters a protein that plays a critical role in the body, a medical condition can result. Some mutations alter a gene's DNA base sequence but do not change the function of the protein made by the gene. One study on the comparison of genes between different species of *Drosophila* suggests that if a mutation does change a protein, this will probably be harmful, with an estimated 70 percent of amino acid polymorphisms having damaging effects, and the remainder being either neutral or weakly beneficial. Studies have shown that only 7 per cent of point mutations in noncoding DNA of yeast are deleterious and 12 per cent in coding DNA are deleterious. The rest of the mutations are either neutral or slightly beneficial. If a mutation is present in a germ cell, it can give rise to offspring that carries the mutation in all of its cells. This is the case in hereditary diseases. In particular, if there is a mutation in a DNA repair gene within a germ cell, humans carrying such germline mutations may have an increased risk of cancer. A list of 34 such germline mutations is given in the article DNA repair-deficiency disorder. An example of one is albinism, a mutation that occurs in the OCA1 or OCA2 gene. Individuals with this disorder are more prone to many types of cancers, other disorders and have impaired vision. On the other hand, a mutation may occur in a somatic cell of an organism. Such mutations will be present in all descendants of this cell within the same organism, and certain mutations can cause the cell to become malignant, and, thus, cause cancer. A DNA damage can cause an error when the DNA is replicated, and this error of replication can cause a gene mutation that, in turn, could cause a genetic disorder. DNA damages are repaired by the DNA repair system of the cell. Each cell has a number of pathways through which enzymes recognize and repair damages in DNA. Because DNA can be damaged in many ways, the process of DNA repair is an important way in which the body protects itself from disease. Once DNA damage has given rise to a mutation, the mutation cannot be repaired. DNA repair pathways can only recognize and act on "abnormal" structures in the DNA. Once a mutation occurs in a gene sequence it then has normal DNA structure and cannot be repaired.

Beneficial Mutations

Although mutations that cause changes in protein sequences can be harmful to an organism, on occasions the effect may be positive in a given environment. In this case, the mutation may enable the mutant organism to withstand particular environmental stresses better than wild-type organisms, or reproduce more quickly. In these cases a mutation will tend to become more common in a population through natural selection. For example, a specific 32 base pair deletion in human CCR5 (CCR5-Δ32) confers HIV resistance to homozygotes and delays AIDS onset in heterozygotes. One possible explanation of the etiology of the relatively high frequency of CCR5-Δ32 in the European population is that it conferred resistance to the bubonic plague in mid-14th century Europe. People with this mutation were more likely to survive infection; thus its frequency in the population increased. This theory could explain why this mutation is not found in Southern Africa, which remained untouched by bubonic plague. A newer theory suggests that the selective pressure on the CCR5 Delta 32 mutation was caused by smallpox instead of the bubonic plague. An example of a harmful mutation is sickle-cell disease, a blood disorder in which the body produces an abnormal type of the oxygen-carrying substance hemoglobin in the red blood cells. One-third of all indigenous inhabitants of Sub-Saharan Africa carry the gene,

because, in areas where malaria is common, there is a survival value in carrying only a single sickle-cell gene (sickle cell trait). Those with only one of the two alleles of the sickle-cell disease are more resistant to malaria, since the infestation of the malaria *Plasmodium* is halted by the sickling of the cells that it infests.

Prion Mutations

Prions are proteins and do not contain genetic material. However, prion replication has been shown to be subject to mutation and natural selection just like other forms of replication. The human gene PRNP codes for the major prion protein, PrP, and is subject to mutations that can give rise to disease-causing prions.

Somatic Mutations

A change in the genetic structure that is not inherited from a parent, and also not passed to offspring, is called a somatic mutation. *Somatic mutations are not inherited because they do not affect the germline. These types of mutations are usually prompted by environmental causes, such as ultraviolet radiation or any exposure to certain harmful chemicals, and can cause diseases including cancer.* With plants, some somatic mutations can be propagated without the need for seed production, for example, by grafting and stem cuttings. These type of mutation have led to new types of fruits, such as the “Delicious” apple and the “Washington” navel orange. Human and mouse somatic cells have a mutation rate more than ten times higher than the germline mutation rate for both species; mice have a higher rate of both somatic and germline mutations per cell division than humans. The disparity in mutation rate between the germline and somatic tissues likely reflects the greater importance of genome maintenance in the germline than in the soma.

Amorphic Mutations

An amorph, a term utilized by Muller in 1932, is a mutated allele, which has lost the ability of the parent (whether wild type or any other type) allele to encode any functional protein. An amorphic mutation may be caused by the replacement of an amino acid that deactivates an enzyme or by the deletion of part of a gene that produces the enzyme. Cells with heterozygous mutations (one good copy of gene and one mutated copy) may function normally with the unmutated copy until the good copy has been spontaneously somatically mutated. This kind of mutation happens all the time in living organisms, but it is difficult to measure the rate. Measuring this rate is important in predicting the rate at which people may develop cancer. Point mutations may arise from spontaneous mutations that occur during DNA replication. The rate of mutation may be increased by mutagens. Mutagens can be physical, such as radiation from UV rays, X-rays or extreme heat, or chemical (molecules that misplace base pairs or disrupt the helical shape of DNA). Mutagens associated with cancers are often studied to learn about cancer and its prevention.

Hypomorphic and Hypermorphic Mutations

A hypomorphic mutation is a replacement of amino acids that would hinder enzyme activity, which would reduce the enzyme level but not to the point of complete loss. Usually, hypomorphic mutations are recessive, but haploinsufficiency causes some alleles to be dominant. A hypermorphic mutation changes the regulation of the gene and causes it to overproduce the gene produce causing a greater than normal enzyme levels. These type of alleles are dominant gain of function type of alleles.

NATURAL SELECTION AND EVOLUTION

Mutations alter an organism’s genotype and occasionally this causes different phenotypes to appear. Most mutations have little effect on an organism’s phenotype, health, or reproductive fitness. Mutations that do have

an effect are usually detrimental, but occasionally some can be beneficial. Studies in the fly *Drosophila melanogaster* suggest that if a mutation changes a protein produced by a gene, about 70 percent of these mutations will be harmful with the remainder being either neutral or weakly beneficial. Population genetics studies the distribution of genetic differences within populations and how these distributions change over time. Changes in the frequency of an allele in a population are mainly influenced by natural selection, where a given allele provides a selective or reproductive advantage to the organism, as well as other factors such as mutation, genetic drift, genetic hitchhiking, artificial selection and migration. Over many generations, the genomes of organisms can change significantly, resulting in evolution. In the process called adaptation, selection for beneficial mutations can cause a species to evolve into forms better able to survive in their environment. New species are formed through the process of speciation, often caused by geographical separations that prevent populations from exchanging genes with each other. By comparing the homology between different species' genomes, it is possible to calculate the evolutionary distance between them and when they may have diverged. Genetic comparisons are generally considered a more accurate method of characterizing the relatedness between species than the comparison of phenotypic characteristics. The evolutionary distances between species can be used to form evolutionary trees; these trees represent the common descent and divergence of species over time, although they do not show the transfer of genetic material between unrelated species (known as horizontal gene transfer and most common in bacteria).

MODEL ORGANISMS

Although geneticists originally studied inheritance in a wide range of organisms, researchers began to specialize in studying the genetics of a particular subset of organisms. The fact that significant research already existed for a given organism would encourage new researchers to choose it for further study, and so eventually a few model organisms became the basis for most genetics research. Common research topics in model organism genetics include the study of gene regulation and the involvement of genes in development and cancer. Organisms were chosen, in part, for convenience—short generation times and easy genetic manipulation made some organisms popular genetics research tools. Widely used model organisms include the gut bacterium *Escherichia coli*, the plant *Arabidopsis thaliana*, baker's yeast (*Saccharomyces cerevisiae*), the nematode *Caenorhabditis elegans*, the common fruit fly (*Drosophila melanogaster*), and the common house mouse (*Mus musculus*).

MEDICINE

Medical genetics seeks to understand how genetic variation relates to human health and disease. When searching for an unknown gene that may be involved in a disease, researchers commonly use genetic linkage and genetic pedigree charts to find the location on the genome associated with the disease. At the population level, researchers take advantage of Mendelian randomization to look for locations in the genome that are associated with diseases, a method especially useful for multigenic traits not clearly defined by a single gene. Once a candidate gene is found, further research is often done on the corresponding (or homologous) genes of model organisms. In addition to studying genetic diseases, the increased availability of genotyping methods has led to the field of pharmacogenetics: the study of how genotype can affect drug responses. Individuals differ in their inherited tendency to develop cancer, and cancer is a genetic disease. The process of cancer development in the body is a combination of events. Mutations occasionally occur within cells in the body as they divide. Although these mutations will not be inherited by any offspring, they can affect the behaviour of cells, sometimes causing them to grow and divide more frequently. There are biological mechanisms that attempt to stop this process; signals are given to inappropriately dividing cells that should trigger cell death, but sometimes additional mutations occur that cause cells to ignore these messages. An internal process of natural selection occurs within the body and eventually mutations accumulate within cells to promote their own growth, creating a cancerous

tumor that grows and invades various tissues of the body. Normally, a cell divides only in response to signals called growth factors and stops growing once in contact with surrounding cells and in response to growth-inhibitory signals. It usually then divides a limited number of times and dies, staying within the epithelium where it is unable to migrate to other organs. To become a cancer cell, a cell has to accumulate mutations in a number of genes (three to seven) that allow it to bypass this regulation: it no longer needs growth factors to divide, continues growing when making contact to neighbour cells, ignores inhibitory signals, keeps growing indefinitely and is immortal, escapes from the epithelium and ultimately may be able to escape from the primary tumor, cross the endothelium of a blood vessel, be transported by the bloodstream and colonize a new organ, forming deadly metastasis. Although there are some genetic predispositions in a small fraction of cancers, the major fraction is due to a set of new genetic mutations that originally appear and accumulate in one or a small number of cells that will divide to form the tumor and are not transmitted to the progeny (somatic mutations). The most frequent mutations are a loss of function of p53 protein, a tumor suppressor, or in the p53 pathway, and gain of function mutations in the Ras proteins, or in other oncogenes.

RESEARCH METHODS

DNA can be manipulated in the laboratory. Restriction enzymes are commonly used enzymes that cut DNA at specific sequences, producing predictable fragments of DNA. DNA fragments can be visualized through use of gel electrophoresis, which separates fragments according to their length. The use of ligation enzymes allows DNA fragments to be connected. By binding (“ligating”) fragments of DNA together from different sources, researchers can create recombinant DNA, the DNA often associated with genetically modified organisms. Recombinant DNA is commonly used in the context of plasmids: short circular DNA molecules with a few genes on them. In the process known as molecular cloning, researchers can amplify the DNA fragments by inserting plasmids into bacteria and then culturing them on plates of agar (to isolate clones of bacteria cells – “cloning” can also refer to the various means of creating cloned (“clonal”) organisms). DNA can also be amplified using a procedure called the polymerase chain reaction (PCR). By using specific short sequences of DNA, PCR can isolate and exponentially amplify a targeted region of DNA. Because it can amplify from extremely small amounts of DNA, PCR is also often used to detect the presence of specific DNA sequences.

DNA SEQUENCING AND GENOMICS

DNA sequencing, one of the most fundamental technologies developed to study genetics, allows researchers to determine the sequence of nucleotides in DNA fragments. The technique of chain-termination sequencing, developed in 1977 by a team led by Frederick Sanger, is still routinely used to sequence DNA fragments. Using this technology, researchers have been able to study the molecular sequences associated with many human diseases. As sequencing has become less expensive, researchers have sequenced the genomes of many organisms using a process called genome assembly, which utilizes computational tools to stitch together sequences from many different fragments. These technologies were used to sequence the human genome in the Human Genome Project completed in 2003. New high-throughput sequencing technologies are dramatically lowering the cost of DNA sequencing, with many researchers hoping to bring the cost of resequencing a human genome down to a thousand dollars. Next-generation sequencing (or high-throughput sequencing) came about due to the ever-increasing demand for low-cost sequencing. These sequencing technologies allow the production of potentially millions of sequences concurrently. The large amount of sequence data available has created the field of genomics, research that uses computational tools to search for and analyze patterns in the full genomes of organisms. Genomics can also be considered a subfield of bioinformatics, which uses computational approaches to analyze large sets of biological data. A common problem to these fields of research is how to manage and share data that deals with human subject and personally identifiable information.

SOCIETY AND CULTURE

On 19 March 2015, a group of leading biologists urged a worldwide ban on clinical use of methods, particularly the use of CRISPR and zinc finger, to edit the human genome in a way that can be inherited. In April 2015, Chinese researchers reported results of basic research to edit the DNA of non-viable human embryos using CRISPR.

Human Microbiota

The human microbiota is the aggregate of microorganisms that resides on or within any of a number of human tissues and biofluids, including the skin, mammary glands, placenta, seminal fluid, uterus, ovarian follicles, lung, saliva, oral mucosa, conjunctiva, biliary and gastrointestinal tracts. They include bacteria, archaea, fungi, protists and viruses. Though micro-animals can also live on the human body, they are typically excluded from this definition. The human microbiome refers specifically to the collective genomes of resident microorganisms. Humans are colonized by many microorganisms; the traditional estimate is that the average human body is inhabited by ten times as many non-human cells as human cells, but more recent estimates have lowered that ratio to 3:1 or even to approximately the same number. Some microorganisms that colonize humans are commensal, meaning they co-exist without harming humans; others have a mutualistic relationship with their human hosts. Conversely, some non-pathogenic microorganisms can harm human hosts via the metabolites they produce, like trimethylamine. Certain microorganisms perform tasks that are known to be useful to the human host but the role of most of them is not well understood. Those that are expected to be present, and that under normal circumstances do not cause disease, are sometimes deemed *normal flora* or *normal microbiota*. The Human Microbiome Project took on the project of sequencing the genome of the human microbiota, focusing particularly on the microbiota that normally inhabit the skin, mouth, nose, digestive tract, and vagina. It reached a milestone in 2012 when it published its initial results.

TERMINOLOGY

Though widely known as *flora ormicroflora*, this is a misnomer in technical terms, since the word root *flora* pertains to plants, and *biota* refers to the total collection of organisms in a particular ecosystem. Recently, the more appropriate term *microbiota* is applied, though its use has not eclipsed the entrenched use and recognition of *flora* with regard to bacteria and other microorganisms. Both terms are being used in different literature.

RELATIVE NUMBERS

As of 2014, it was often reported in popular media and in the scientific literature that there are about 10 times as many microbial cells in the human body as there are human cells; this figure was based on estimates that the human microbiome includes around 100 trillion bacterial cells and that an adult human typically has around 10 trillion human cells. In 2014, the American Academy of Microbiology published a FAQ that emphasized that the number of microbial cells and the number of human cells are both estimates, and noted that recent research had arrived at a new estimate of the number of human cells – approximately 37.2 trillion, meaning that the ratio of microbial-to-human cells, if the original estimate of 100 trillion bacterial cells is correct, is closer to 3:1. In 2016, another group published a new estimate of the ratio being roughly 1:1 (1.3:1, with “an uncertainty of 25 per cent and a variation of 53 per cent over the population of standard 70-kg males”).

STUDY

The problem of elucidating the human microbiome is essentially identifying the members of a microbial community which includes bacteria, eukaryotes, and viruses. This is done primarily using DNA-based studies, though RNA, protein and metabolite based studies are also performed. DNA-based microbiome studies typically can be categorized as either targeted amplicon studies or more recently shotgun metagenomic studies. The former focuses on specific known marker genes and is primarily informative taxonomically, while the latter is an entire metagenomic approach which can also be used to study the functional potential of the community. One of the challenges that is present in human microbiome studies, but not in other metagenomic studies is to avoid including the host DNA in the study. Aside from simply elucidating the composition of the human microbiome, one of the major questions involving the human microbiome is whether there is a “core”, that is, whether there is a subset of the community that is shared among most humans. If there is a core, then it would be possible to associate certain community compositions with disease states, which is one of the goals of the Human Microbiome Project. It is known that the human microbiome (such as the gut microbiota) is highly variable both within a single subject and among different individuals, a phenomenon which is also observed in mice. On 13 June 2012, a major milestone of the Human Microbiome Project (HMP) was announced by the NIH director Francis Collins.

The announcement was accompanied with a series of coordinated articles published in *Nature* and several journals in the Public Library of Science (PLOS) on the same day. By mapping the normal microbial make-up of healthy humans using genome sequencing techniques, the researchers of the HMP have created a reference database and the boundaries of normal microbial variation in humans. From 242 healthy U.S. volunteers, more than 5,000 samples were collected from tissues from 15 (men) to 18 (women) body sites such as mouth, nose, skin, lower intestine (stool), and vagina. All the DNA, human and microbial, were analyzed with DNA sequencing machines. The microbial genome data were extracted by identifying the bacterial specific ribosomal RNA, 16S rRNA. The researchers calculated that more than 10,000 microbial species occupy the human ecosystem and they have identified 81 – 99 per cent of the genera.

TYPES

Bacteria

Populations of microbes (such as bacteria and yeasts) inhabit the skin and mucosal surfaces in various parts of the body. Their role forms part of normal, healthy human physiology, however if microbe numbers grow beyond their typical ranges (often due to a compromised immune system) or if microbes populate (such as through poor hygiene or injury) areas of the body normally not colonized or sterile (such as the blood, or the lower respiratory tract, or the abdominal cavity), disease can result (causing, respectively, bacteremia/sepsis, pneumonia, and peritonitis). The Human Microbiome Project found that individuals host thousands of bacterial types, different body sites having their own distinctive communities. Skin and vaginal sites showed smaller diversity than the mouth and gut, these showing the greatest richness. The bacterial makeup for a given site on a body varies from person to person, not only in type, but also in abundance. Bacteria of the same species found throughout the mouth are of multiple subtypes, preferring to inhabit distinctly different locations in the mouth. Even the enterotypes in the human gut, previously thought to be well understood, are from a broad spectrum of communities with blurred taxon boundaries. It is estimated that 500 to 1,000 species of bacteria live in the human gut but belong to just a few phyla: Firmicutes and Bacteroidetes dominate but there are also Proteobacteria, Verrucomicrobia, Actinobacteria, Fusobacteria and Cyanobacteria. A number of types of bacteria, such as *Actinomyces viscosus* and *A. naeslundii*, live in the mouth, where they are part of a sticky substance called

plaque. If this is not removed by brushing, it hardens into calculus (also called tartar). The same bacteria also secrete acids that dissolve tooth enamel, causing tooth decay. The vaginal microflora consist mostly of various lactobacillus species. It was long thought that the most common of these species was *Lactobacillus acidophilus*, but it has later been shown that *L. iners* is in fact most common, followed by *L. crispatus*. Other lactobacilli found in the vagina are *L. jensenii*, *L. delbruekii* and *L. gasseri*. Disturbance of the vaginal flora can lead to infections such as bacterial vaginosis or candidiasis (“yeast infection”).

ARCHAEA

Archaea are present in the human gut, but, in contrast to the enormous variety of bacteria in this organ, the numbers of archaeal species are much more limited. The dominant group are the methanogens, particularly *Methanobrevibacter smithii* and *Methanosphaera stadtmanae*. However, colonization by methanogens is variable, and only about 50 per cent of humans have easily detectable populations of these organisms. As of 2007, no clear examples of archaeal pathogens were known, although a relationship has been proposed between the presence of some methanogens and human periodontal disease.

FUNGI

Fungi, in particular yeasts, are present in the human gut. The best-studied of these are *Candida* species due to their ability to become pathogenic in immunocompromised and even in healthy hosts. Yeasts are also present on the skin, such as *Malassezia* species, where they consume oils secreted from the sebaceous glands.

VIRUSES

Viruses, especially bacterial viruses (bacteriophages), colonize various body sites. These colonized sites include the skin, gut, lungs, and oral cavity. Virus communities have been associated with some diseases, and do not simply reflect the bacterial communities.

ANATOMICAL AREAS

Skin

The term skin flora (also commonly referred to as skin microbiota) refers to the microorganisms which reside on the skin, typically human skin. Many of them are bacteria of which there are around 1000 species upon human skin from nineteen phyla. Most are found in the superficial layers of the epidermis and the upper parts of hair follicles. Skin flora is usually non-pathogenic, and either commensal (are not harmful to their host) or mutualistic (offer a benefit). The benefits bacteria can offer include preventing transient pathogenic organisms from colonizing the skin surface, either by competing for nutrients, secreting chemicals against them, or stimulating the skin’s immune system. However, resident microbes can cause skin diseases and enter the blood system, creating life-threatening diseases, particularly in immunosuppressed people. A major non-human skin flora is *Batrachochytrium dendrobatidis*, a chytrid and non-hyphal zoosporic fungus that causes chytridiomycosis, an infectious disease thought to be responsible for the decline in amphibian populations.

Species Variety

Bacteria

The estimate of the number of species present on skin bacteria has been radically changed by the use of 16S ribosomal RNA to identify bacterial species present on skin samples direct from their genetic material. Previously

such identification had depended upon microbiological culture upon which many varieties of bacteria did not grow and so were hidden to science. *Staphylococcus epidermidis* and *Staphylococcus aureus* were thought from cultural based research to be dominant. However 16S ribosomal RNA research finds that while common, these species make up only 5 per cent of skin bacteria. However, skin variety provides a rich and diverse habitat for bacteria. Most come from four phyla: Actinobacteria (51.8 per cent), Firmicutes (24.4 per cent), Proteobacteria (16.5 per cent), and Bacteroidetes (6.3 per cent).

There are three main ecological areas: sebaceous, moist, and dry. *Propionibacteria* and *Staphylococci* species were the main species in sebaceous areas. In moist places on the body *Corynebacteria* together with *Staphylococci* dominate. In dry areas, there is a mixture of species but b-Proteobacteria and Flavobacteriales are dominant. Ecologically, sebaceous areas had greater species richness than moist and dry one. The areas with least similarity between people in species were the spaces between fingers, the spaces between toes, axillae, and umbilical cord stump. Most similarly were beside the nostril, nares (inside the nostril), and on the back.

Table. Frequency of the best studied skin microbes.

Organism	Observations
<i>Staphylococcus epidermidis</i>	Common, occasionally pathogenic
<i>Staphylococcus aureus</i>	Infrequent, usually pathogenic
<i>Staphylococcus warneri</i>	Infrequent, occasionally pathogenic
<i>Streptococcus pyogenes</i>	Infrequent, usually pathogenic
<i>Streptococcus mitis</i>	Frequent, occasionally pathogenic
<i>Propionibacterium acnes</i>	Frequent, occasionally pathogenic
<i>Corynebacterium</i> spp.	Frequent, occasionally pathogenic
<i>Acinetobacter johnsonii</i>	Frequent, occasionally pathogenic
<i>Pseudomonas aeruginosa</i>	Infrequent, occasionally pathogenic

Fungal

A study of the area between toes in 100 young adults found 14 different genera of fungi. These include yeasts such as *Candida albicans*, *Rhodotorula rubra*, *Torulopsis* and *Trichosporon cutaneum*, dermatophytes (skin living fungi) such as *Microsporum gypseum*, and *Trichophyton rubrum* and nondermatophyte fungi (opportunistic fungi that can live in skin) such as *Rhizopus stolonifer*, *Trichosporon cutaneum*, *Fusarium*, *Scopulariopsis brevicaulis*, *Curvularia*, *Alternaria alternata*, *Paecilomyces*, *Aspergillus flavus* and *Penicillium* species. A study by the National Human Genome Research Institute in Bethesda, Maryland, researched the DNA of human skin fungi at 14 different locations on the body. These were the ear canal, between the eyebrows, the back of the head, behind the ear, the heel, toenails, between the toes, forearm, back, groin, nostrils, chest, palm, and the crook of the elbow. The study showed a large fungal diversity across the body, the richest habitat being the heel, which hosts about 80 species of fungi. By way of contrast, there are some 60 species in toenail clippings and 40 between the toes. Other rich areas are the palm, forearm and inside the elbow, with from 18 to 32 species. The head and the trunk hosted between 2 and 10 each.

Umbilical Microbiome

The umbilicus, or navel, is an area of the body that is rarely exposed to UV light, soaps, or bodily secretions (the navel does not produce any secretions or oils) and because it is an almost undisturbed community of bacteria it is an excellent part of the skin microbiome to study. The navel, or umbilicus is a moist microbiome of the body (with high humidity and temperatures), that contains a large amount of bacteria, especially bacteria that favors moist conditions such as *Corynebacterium* and *Staphylococcus*. The Belly Button Biodiversity Project began at North Carolina State University in early 2011 with two initial groups of 35 and 25 volunteers. Volunteers were given sterile cotton swabs and were asked to insert the cotton swabs into their navels, to turn

the cotton swab around three times and then return the cotton swab to the researchers in a vial that contained a 0.5 ml 10 per cent phosphate saline buffer. Researchers at North Carolina State University, led by Jiri Hulcr, then grew the samples in a culture until the bacterial colonies were large enough to be photographed and then these pictures were posted on the Belly Button Biodiversity Project's web site (volunteers were given sample numbers so that they could view their own samples online). These samples then were analyzed using 16S rDNA libraries so that strains that did not grow well in cultures could be identified. The researchers at North Carolina State University discovered that while it was difficult to predict every strain of bacteria in the microbiome of the navel that they could predict which strains would be prevalent and which strains of bacteria would be quite rare in the microbiome. It was found that the navel microbiomes only contained a few prevalent types of bacteria (*Staphylococcus*, *Corynebacterium*, Actinobacteria, Clostridiales, and Bacilli) and many different types of rare bacteria. Other types of rare organisms were discovered inside the navels of the volunteers including three types of Archaea (an organism that usually lives in only extreme environments) and two of the three types of Archaea were found in one volunteer who claimed not to have bathed or showered for many years.

Staphylococcus and *Corynebacterium* were among the most common types of bacteria found in the navels of this project's volunteers and these types of bacteria have been found to be the most common types of bacteria found on the human skin in larger studies of the skin microbiome (of which the Belly Button Biodiversity Project is a part). (In these larger studies it has been found that females generally have more *Staphylococcus* living in their skin microbiomes (usually *Staphylococcus epidermidis*) and that men have more *Corynebacterium* living in their skin microbiomes.) According to the Belly Button Biodiversity Project at North Carolina State University, there are two types of microorganisms found in the navel and surrounding areas. Transient bacteria (bacteria that does not reproduce) forms the majority of the organisms found in the navel, and an estimated 1400 various strains were found in 95 per cent of participants of the study. The Belly Button Biodiversity Project is ongoing and has now taken swabs from over 500 people. The project was designed with the aim of countering that misconception that bacteria are always harmful to humans and that humans are at war with bacteria. In actuality, most strains of bacteria are harmless if not beneficial for the human body. Another of the project's goals is to foster public interest in microbiology. Working in concert with the Human Microbiome Project, the Belly Button Biodiversity Project also studies the connections between human microbiomes and the factors of age, sex, ethnicity, location and overall health.

Relationship to Host

Skin microflora can be commensals, mutualistic or pathogens. Often they can be all three depending upon the strength of the person's immune system. Research upon the immune system in the gut and lungs has shown that microflora aids immunity development: however such research has only started upon whether this is the case with the skin. *Pseudomonas aeruginosa* is an example of a mutualistic bacterium that can turn into a pathogen and cause disease: if it gains entry into the blood system it can result in infections in bone, joint, gastrointestinal, and respiratory systems. It can also cause dermatitis. However, *Pseudomonas aeruginosa* produces antimicrobial substances such as pseudomonic acid (that are exploited commercially such as Mupirocin).

This works against staphylococcal and streptococcal infections. *Pseudomonas aeruginosa* also produces substances that inhibit the growth of fungus species such as *Candida krusei*, *Candida albicans*, *Torulopsis glabrata*, *Saccharomyces cerevisiae* and *Aspergillus fumigatus*. It can also inhibit the growth of *Helicobacter pylori*. So important is its antimicrobial actions that it has been noted that "removing *P. aeruginosa* from the skin, through use of oral or topical antibiotics, may inversely allow for aberrant yeast colonization and infection." Another aspect of bacteria is the generation of body odour. Sweat is odorless however several bacteria may consume it and create byproducts which may be considered putrid by man (as in contrast to flies, for example, that may find them attractive/appealing). Several examples are:

- *Propionibacteria* in adolescent and adult sebaceous glands can turn its amino acids into propionic acid.
- *Staphylococcus epidermidis* creates body odour by breaking sweat into isovaleric acid (3-methyl butanoic acid).
- *Bacillus subtilis* creates strong foot odour.

Skin Defenses

Antimicrobial Peptides

The skin creates antimicrobial peptides such as cathelicidins that control the proliferation of skin microbes. Cathelicidins not only reduce microbe numbers directly but also cause the secretion of cytokine release which induces inflammation, angiogenesis, and reepithelialization. Conditions such as atopic dermatitis have been linked to the suppression in cathelicidin production. In rosacea abnormal processing of cathelicidin cause inflammation. Psoriasis has been linked to self-DNA created from cathelicidin peptides that causes autoinflammation. A major factor controlling cathelicidin is vitamin D₃.

Acidity

The superficial layers of the skin are naturally acidic (pH 4-4.5) due to lactic acid in sweat and produced by skin bacteria. At this pH mutualistic flora such as *Staphylococci*, *Micrococci*, *Corynebacterium* and *Propionibacteria* grow but not transient bacteria such as Gram negative bacteria like *Escherichia* and *Pseudomonas* or Gram positive ones such as *Staphylococcus aureus*. Another factor affecting the growth of pathological bacteria is that the antimicrobial substances secreted by the skin are enhanced in acidic conditions. In alkaline conditions, bacteria cease to be attached to the skin and are more readily shed. It has been observed that the skin also swells under alkaline conditions and opens up allowing move to the surface.

Immune System

If activated, the immune system in the skin produces cell-mediated immunity against microbes such as dermatophytes (skin fungi). One reaction is to increase stratum corneum turnover and so shed the fungus from the skin surface. Skin fungi such as *Trichophyton rubrum* have evolved to create substances that limit the immune response to them. The shedding of skin is a general means to control the buildup of flora upon the skin surface.

Skin Diseases

Microorganisms play a role in noninfectious skin diseases such as atopic dermatitis, rosacea, psoriasis, and acne. Damaged skin can cause nonpathogenic bacteria to become pathogenic. The diversity of species on the skin is related to later development of dermatitis.

Acne Vulgaris

Acne vulgaris is a common skin condition characterised by excessive sebum production by the pilosebaceous unit and inflammation of the skin. Affected areas are typically colonised by *Propionibacterium acnes*; a member of the commensal microbiota even in those without acne. High populations of *P. acnes* are linked to acne vulgaris although only certain strains are strongly associated with acne while others with healthy skin. The relative population of *P. acnes* is similar between those with acne and those without. Current treatment

includes topical and systemic antibacterial drugs which result in decreased *P. acnes* colonisation and/or activity. Potential probiotic treatment includes the use of *Staphylococcus epidermidis* to inhibit *P. acnes* growth. *S. epidermidis* produces succinic acid which has been shown to inhibit *P. acnes* growth. *Lactobacillus plantarum* has also been shown to act as an anti-inflammatory and improve antimicrobial properties of the skin when applied topically. It was also shown to be effective in reducing acne lesion size.

Atopic Dermatitis

Individuals with atopic dermatitis have shown an increase in populations of *Staphylococcus aureus* in both lesional and nonlesional skin. Atopic dermatitis flares are associated with low bacterial diversity due to colonisation by *S. aureus* and following standard treatment, bacterial diversity has been seen to increase. Current treatments include combinations of topical or systemic antibiotics, corticosteroids, and diluted bleach baths. Potential probiotic treatments include using the commensal skin bacteria, *S. epidermidis*, to inhibit *S. aureus* growth. During atopic dermatitis flares, population levels of *S. epidermidis* has been shown to increase as an attempt to control *S. aureus* populations. Low gut microbial diversity in babies has been associated with an increased risk of atopic dermatitis. Infants with atopic eczema have low levels of *Bacteroides* and high levels of Firmicutes. *Bacteroides* have anti-inflammatory properties which are essential against dermatitis.

Psoriasis Vulgaris

Psoriasis vulgaris typically affects drier skin sites such as elbows and knees. Dry areas of the skin tend to have high microbial diversity and fewer populations than sebaceous sites. A study using swab sampling techniques show areas rich in Firmicutes (mainly *Streptococcus* and *Staphylococcus*) and Actinobacteria (mainly *Corynebacterium* and *Propionibacterium*) are associated with psoriasis. While another study using biopsies associate increased levels of Firmicutes and Actinobacteria with healthy skin. However most studies show that individuals affected by psoriasis have a lower microbial diversity in the affected areas. Treatments for psoriasis include topical agents, phototherapy, and systemic agents. Current research on the skin microbiota's role in psoriasis is inconsistent therefore there are no potential probiotic treatments.

Rosacea

Rosacea is typically connected to sebaceous sites of the skin. The skin mite *Demodex folliculorum* produce lipases that allow them to use sebum as a source of food therefore they have a high affinity for sebaceous skin sites. Although it is a part of the commensal skin microbiota, patients affected with rosacea show an increase in *D. folliculorum* compared to healthy individuals, suggesting pathogenicity.

Bacillus oleronius, a *Demodex* associated microbe, is not typically found in the commensal skin microbiota but initiates inflammatory pathways whose starting mechanism is similar to rosacea patients. Populations of *S. epidermidis* have also been isolated from pustules of rosacea patients. However it is possible that they were moved by *Demodex* to areas that favour growth as *Demodex* has shown to transport bacteria around the face. Current treatments include topical and oral antibiotics and laser therapy. As current research has yet to show a clear mechanism for *Demodex* influence in rosacea, there are no potential probiotic treatments.

Clinical

Infected Devices

Skin microbes are a potential source of infected medical devices such as catheters.

Hygiene

It is important to note that the human skin is host to numerous bacterial and fungal species, some of which are known to be harmful, some known to be beneficial and the vast majority unresearched. The use of bactericidal and fungicidal soaps will inevitably lead to bacterial and fungal populations which are resistant to the chemicals employed.

Contagion

Skin flora do not readily pass between people: 30 seconds of moderate friction and dry hand contact results in a transfer of only 0.07 per cent of natural hand flora from naked with a greater percentage from gloves.

Removal

The most effective (60 to 80 per cent reduction) antimicrobial washing is with ethanol, isopropanol, and n-propanol. Viruses are most affected by high (95 per cent) concentrations of ethanol, while bacteria are more affected by n-propanol. Unmedicated soaps are not very effective as illustrated by the following data. Health care workers washed their hands once in nonmedicated liquid soap for 30 seconds. The students/technicians for 20 times.

Table. Skin flora upon two hospital groups in colony-forming units per mL.

Group and hand skin condition	Unwashed	Washed
Health care workers healthy	3.47	3.15
Health care workers damaged	3.33	3.29
Students/technicians healthy	4.39	3.54
Students/technicians damaged	4.58	4.43

An important use of hand washing is to prevent the transmission of antibiotic resistant skin flora that cause hospital-acquired infections such as Methicillin-resistant *Staphylococcus aureus*. While such flora have become antibiotic resistant due to antibiotics there is no evidence that recommended antiseptics or disinfectants selects for antibiotic-resistant organisms when used in hand washing. However, many strains of organisms are resistant to some of the substances used in antibacterial soaps such as Triclosan. One survey of bar soaps in dentist clinics found they all had their own flora and on average from two to five different genera of microorganisms with those used most more likely to have more species varieties. Another survey of bar soaps in public toilets found even more flora. Another study found that very dry soaps are not infected while all are that rest in pools of water. However, research upon soap that was specially infected found that soap flora do not transmit to the hands.

Damaged Skin

Washing skin repeatedly can damage the protective external layer and cause transepidermal loss of water. This can be seen in roughness characterized by scaling and dryness, itchiness, dermatitis provoked by microorganisms and allergens penetrating the corneal layer and redness. Wearing gloves can cause further problems since it produces a humid environment favoring the growth of microbes and also contains irritants such as latex and talcum powder.

Hand washing can damage skin because the stratum corneum top layer of skin consists of 15 to 20 layers of keratin disks, corneocytes, each of which is each surrounded by a thin film of skin lipids which can be removed by alcohols and detergents. Damaged skin defined by extensive cracking of skin surface, widespread reddening or occasional bleeding has also been found to be more frequently colonized by *Staphylococcus hominis* and

these were more likely to be methicillin resistant. Though not related to greater antibiotic resistance, damaged skin was also more likely to be colonized by *Staphylococcus aureus*, gram-negative bacteria, *Enterococci* and *Candida*.

Comparison with other Flora

The skin flora is different from that of the gut which is predominantly Firmicutes and Bacteroidetes. There is also low level of variation between people that is not found in gut studies. Both gut and skin flora however lack the diversity found in soil flora.

CONJUNCTIVA

A small number of bacteria and fungi are normally present in the conjunctiva. Classes of bacteria include Gram-positive cocci (*e.g.*, *Staphylococcus* and *Streptococcus*) and Gram-negative rods and cocci (*e.g.*, *Haemophilus* and *Neisseria*) are present. Fungal genera include *Candida*, *Aspergillus*, and *Penicillium*. The lachrymal glands continuously secrete, keeping the conjunctiva moist, while intermittent blinking lubricates the conjunctiva and washes away foreign material. Tears contain bactericides such as lysozyme, so that microorganisms have difficulty in surviving the lysozyme and settling on the epithelial surfaces.

GUT

Gut flora, or gut microbiota, or gastrointestinal microbiota, is the complex community of microorganisms that live in the digestive tracts of humans and other animals, including insects. The gut metagenome is the aggregate of all the genomes of gut microbiota. The gut is one niche that human microbiota inhabit. In humans, the gut microbiota has the largest numbers of bacteria and the greatest number of species compared to other areas of the body. In humans, the gut flora is established at one to two years after birth, and by that time the intestinal epithelium and the intestinal mucosal barrier that it secretes have co-developed in a way that is tolerant to, and even supportive of, the gut flora and that also provides a barrier to pathogenic organisms. The relationship between some gut flora and humans is not merely commensal (a non-harmful coexistence), but rather a mutualistic relationship. Some human gut microorganisms benefit the host by fermenting dietary fibre into short-chain fatty acids (SCFAs), such as acetic acid and butyric acid, which are then absorbed by the host. Intestinal bacteria also play a role in synthesizing vitamin B and vitamin K as well as metabolizing bile acids, sterols, and xenobiotics. The systemic importance of the SCFAs and other compounds they produce are like hormones and the gut flora itself appears to function like an endocrine organ, and dysregulation of the gut flora has been correlated with a host of inflammatory and autoimmune conditions. The composition of human gut microbiota changes over time, when the diet changes, and as overall health changes. A systematic review from 2016 examined the preclinical and small human trials that have been conducted with certain commercially available strains of probiotic bacteria and identified those that had the most potential to be useful for certain central nervous system disorders.

Classifications

The microbial composition of the gut microbiota varies across the digestive tract. In the stomach and small intestine, relatively few species of bacteria are generally present. The colon, in contrast, contains a densely-populated microbial ecosystem with up to 10 cells per gram of intestinal content. These bacteria represent between 300 and 1000 different species. However, 99 per cent of the bacteria come from about 30 or 40 species. As a consequence of their abundance in the intestine, bacteria also make up to 60 per cent of the dry mass of feces. Fungi, protists, archaea, and viruses are also present in the gut flora, but less is known about their activities. Over 99 per cent of the bacteria in the gut are anaerobes, but in the cecum, aerobic bacteria reach high

densities. It is estimated that these gut flora have around a hundred times as many genes in total as there are in the human genome. Many species in the gut have not been studied outside of their hosts because most cannot be cultured. While there are a small number of core species of microbes shared by most individuals, populations of microbes can vary widely among different individuals. Within an individual, microbe populations stay fairly constant over time, even though some alterations may occur with changes in lifestyle, diet and age. The Human Microbiome Project has set out to better describe the microflora of the human gut and other body locations.

The four dominant bacterial phyla in the human gut are Firmicutes, Bacteroidetes, Actinobacteria, and Proteobacteria. Most bacteria belong to the genera *Bacteroides*, *Clostridium*, *Faecalibacterium*, *Eubacterium*, *Ruminococcus*, *Peptococcus*, *Peptostreptococcus*, and *Bifidobacterium*. Other genera, such as *Escherichia* and *Lactobacillus*, are present to a lesser extent. Species from the genus *Bacteroides* alone constitute about 30 per cent of all bacteria in the gut, suggesting that this genus is especially important in the functioning of the host. Fungal genera that have been detected in the gut include *Candida*, *Saccharomyces*, *Aspergillus*, *Penicillium*, *Rhodotorula*, *Trametes*, *Pleospora*, *Sclerotinia*, *Bullera*, and *Galactomyces*, among others. *Rhodotorula* is most frequently found in individuals with inflammatory bowel disease while *Candida* is most frequently found in individuals with hepatitis B cirrhosis and chronic hepatitis B. Archaea constitute another large class of gut flora which are important in the metabolism of the bacterial products of fermentation.

Enterotype

An enterotype is a classification of living organisms based on its bacteriological ecosystem in the human gut microbiome not dictated by age, gender, body weight, or national divisions. There are indications that long-term diet influences enterotype. Three human enterotypes have been proposed, but their value has been questioned.

Flora Composition

Anatomy

Stomach Flora

Due to the high acidity of the stomach, most microorganisms cannot survive there. The main bacterial inhabitants of the stomach include: *Streptococcus*, *Staphylococcus*, *Lactobacillus*, *Peptostreptococcus*, and types of yeast. *Helicobacter pylori* is a Gram-negative spiral organism that establishes on gastric mucosa causing chronic gastritis and peptic ulcer disease and is a carcinogen for gastric cancer.

Intestinal Flora

The small intestine contains a trace amount of microorganisms due to the proximity and influence of the stomach. Gram-positive cocci and rod-shaped bacteria are the predominant microorganisms found in the small intestine. However, in the distal portion of the small intestine alkaline conditions support gram-negative bacteria of the *Enterobacteriaceae*. The bacterial flora of the small intestine aid in a wide range of intestinal functions. The bacterial flora provide regulatory signals that enable the development and utility of the gut. Overgrowth of bacteria in the small intestine can lead to intestinal failure. In addition the large intestine contains the largest bacterial ecosystem in the human body. About 99 per cent of the large intestine and feces flora are made up of obligate anaerobes such as *Bacteroides* and *Bifidobacterium*. Factors that disrupt the microorganism population of the large intestine include antibiotics, stress, and parasites. Bacteria make up most of the flora in the colon and 60 per cent of the dry mass of feces. This fact makes feces an ideal source to test for gut flora for any tests and experiments by extracting the nucleic acid from fecal specimens, and bacterial 16S rRNA gene sequences are generated with bacterial primers. This form of testing is also often preferable to more invasive techniques,

such as biopsies. Somewhere between 300 and 1000 different species live in the gut, with most estimates at about 500. However, it is probable that 99 per cent of the bacteria come from about 30 or 40 species, with *Faecalibacterium prausnitzii* being the most common species in healthy adults. Fungi and protists also make up a part of the gut flora, but less is known about their activities. The virome is mostly bacteriophages. Research suggests that the relationship between gut flora and humans is not merely commensal (a non-harmful coexistence), but rather is a mutualistic, symbiotic relationship. Though people can survive with no gut flora, the microorganisms perform a host of useful functions, such as fermenting unused energy substrates, training the immune system via end products of metabolism like propionate and acetate, preventing growth of harmful species, regulating the development of the gut, producing vitamins for the host (such as biotin and vitamin K), and producing hormones to direct the host to store fats. Extensive modification and imbalances of the gut microbiota and its microbiome or gene collection are associated with obesity. However, in certain conditions, some species are thought to be capable of causing disease by causing infection or increasing cancer risk for the host.

Age

It has been demonstrated that there are common patterns of microbiome composition evolution during life. In general, the diversity of microbiota composition of fecal samples is significantly higher in adults than in children, although interpersonal differences are higher in children than in adults. Much of the maturation of microbiota into an adult-like configuration happens during the three first years of life. As the microbiome composition changes, so does the composition of bacterial proteins produced in the gut. In adult microbiomes, a high prevalence of enzymes involved in fermentation, methanogenesis and the metabolism of arginine, glutamate, aspartate and lysine have been found. In contrast, in infant microbiomes the dominant enzymes are involved in cysteine metabolism and fermentation pathways.

Diet

Studies and statistical analyses have identified the different bacterial genera in gut microbiota and their associations with nutrient intake. Gut microflora is mainly composed of three enterotypes: *Prevotella*, *Bacteroides*, and *Ruminococcus*. There is an association between the concentration of each microbial community and diet. For example, *Prevotella* is related to carbohydrates and simple sugars, while *Bacteroides* is associated with proteins, amino acids, and saturated fats. Specialist microbes that break down mucin, survive on their host's carbohydrate excretions. One enterotype will dominate depending on the diet. Altering the diet will result in a corresponding change in the numbers of species. Malnourished human children have less mature and less diverse gut microbiota than healthy children, and changes in the microbiome associated with nutrient scarcity can in turn be a pathophysiological cause of malnutrition. Malnourished children also typically have more potentially pathogenic gut flora, and more yeast in their mouths and throats. Altering diet may lead to changes in gut microbiota composition and diversity.

Geography

Gut microbiome composition depends on the geographic origin of populations. Variations in a trade-off of *Prevotella*, the representation of the urease gene, and the representation of genes encoding glutamate synthase/degradation or other enzymes involved in amino acids degradation or vitamin biosynthesis show significant differences between populations from the US, Malawi or Amerindian origin. The US population has a high representation of enzymes encoding the degradation of glutamine and enzymes involved in vitamin and lipoic acid biosynthesis; whereas Malawi and Amerindian populations have a high representation of enzymes encoding glutamate synthase and they also have an overrepresentation of α -amylase in their microbiomes. As the US

population has a diet richer in fats than Amerindian or Malawian populations which have a corn-rich diet, the diet is probably the main determinant of the gut bacterial composition. Further studies have indicated a large difference in the composition of microbiota between European and rural African children. The fecal bacteria of children from Florence were compared to that of children from the small rural village of Boulpon in Burkina Faso. The diet of a typical child living in this village is largely lacking in fats and animal proteins and rich in polysaccharides and plant proteins. The fecal bacteria of European children were dominated by *Firmicutes* and showed a marked reduction in biodiversity, while the fecal bacteria of the Boulpon children was dominated by *Bacteroidetes*. The increased biodiversity and different composition of gut flora in African populations may aid in the digestion of normally indigestible plant polysaccharides and also may result in a reduced incidence of non-infectious colonic diseases. On a smaller scale, it has been shown that sharing numerous common environmental exposures in a family is a strong determinant of individual microbiome composition. This effect has no genetic influence and it is consistently observed in culturally different populations.

Acquisition of Gut Flora in Human Infants

The establishment of a gut flora is crucial to the health of an adult, as well the functioning of the gastrointestinal tract. In humans, a gut flora similar to an adult's is formed within one to two years of birth. The traditional view of the gastrointestinal tract of a normal fetus is that it is sterile, although this view has been challenged in the past few years. Multiple lines of evidence have begun to emerge that suggest there may be bacteria in the intrauterine environment. In humans, research has shown that microbial colonization may occur in the fetus with one study showing *Lactobacillus* and *Bifidobacterium* species were present in placental biopsies. Several rodent studies have demonstrated the presence of bacteria in the amniotic fluid and placenta, as well as in the meconium of babies born by sterile cesarean section. In another study, researchers administered a culture of bacteria orally to a pregnant dam, and detected the bacteria in the offspring, likely resulting from transmission between the digestive tract and amniotic fluid via the blood stream. However, researchers caution that the source of these intrauterine bacteria, whether they are alive, and their role, is not yet understood.

During birth and rapidly thereafter, bacteria from the mother and the surrounding environment colonize the infant's gut. The exact sources of bacteria is not fully understood, but may include the birth canal, other people (parents, siblings, hospital workers), breastmilk, food, and the general environment with which the infant interacts. However, as of 2013, it remains unclear whether most colonizing arises from the mother or not. Infants born by caesarean section may also be exposed to their mothers' microflora, but the initial exposure is most likely to be from the surrounding environment such as the air, other infants, and the nursing staff, which serve as vectors for transfer. During the first year of life, the composition of the gut flora is generally simple and changes a great deal with time and is not the same across individuals. The initial bacterial population are generally facultative anaerobic organisms; investigators believe that these initial colonizers decrease the oxygen concentration in the gut, which in turn allows obligately anaerobic bacteria like *Bacteroides*, *Actinobacteria*, and *Firmicutes* to become established and thrive. Breast-fed babies become dominated by bifidobacteria, possibly due to the contents of bifidobacterial growth factors in breast milk, and by the fact that breast milk carries prebiotic components, allowing for healthy bacterial growth. In contrast, the microbiota of formula-fed infants is more diverse, with high numbers of *Enterobacteriaceae*, enterococci, bifidobacteria, *Bacteroides*, and clostridia.

Caesarean section, antibiotics, and formula feeding may alter the gut microbiome composition. Children treated with antibiotics have less stable, and less diverse floral communities. Caesarean sections have been shown to be disruptive to mother-offspring transmission of bacteria, which impacts the overall health of the offspring by raising risks of disease such as celiacs, asthma, and type 1 diabetes. This further evidences the importance of a healthy gut microbiome. Various methods of microbiome restoration are being explored, typically involving exposing the infant to maternal vaginal contents, and oral probiotics.

Functions

When the gut flora first started to be studied, it was thought to have three key roles: directly defending against pathogens, fortifying host defence by its role in developing and maintaining the intestinal epithelium and inducing antibody production there, and metabolizing otherwise indigestible compounds in food; subsequent work discovered its role in training the developing immune system, and yet further work focused on its role in the gut-brain axis.

Direct Inhibition of Pathogens

The gut flora community plays a direct role in defending against pathogens by fully colonizing the space, making use of all available nutrients, and by secreting compounds that kill or inhibit unwelcome organisms that would compete for nutrients with it. Disruption of the gut flora allows competing organisms like *Clostridium difficile* to become established that otherwise are kept in abeyance.

Development of Enteric Protection and Immune System

In humans, a gut flora similar to an adult's is formed within one to two years of birth. As the gut flora gets established, the lining of the intestines – the intestinal epithelium and the intestinal mucosal barrier that it secretes – develop as well, in a way that is tolerant to, and even supportive of, commensurate microorganisms to a certain extent and also provides a barrier to pathogenic ones. Specifically, goblet cells that produce the mucosa proliferate, and the mucosa layer thickens, providing an outside mucosal layer in which “friendly” microorganisms can anchor and feed, and an inner layer that even these organisms cannot penetrate. Additionally, the development of gut-associated lymphoid tissue (GALT), which forms part of the intestinal epithelium and which detects and reacts to pathogens, appears and develops during the time that the gut flora develops and established. The GALT that develops is tolerant to gut flora species, but not to other microorganisms. GALT also normally becomes tolerant to food to which the infant is exposed, as well as digestive products of food, and gut flora's metabolites produced from food.

The human immune system creates cytokines that can drive the immune system to produce inflammation in order to protect itself, and that can tamp down the immune response to maintain homeostasis and allow healing after insult or injury. Different bacterial species that appear in gut flora have been shown to be able to drive the immune system to create cytokines selectively; for example *Bacteroides fragilis* and some *Clostridia* species appear to drive an anti-inflammatory response, while some segmented filamentous bacteria drive the production of inflammatory cytokines. Gut flora can also regulate the production of antibodies by the immune system. One function of this regulation is to cause B cells to class switch to IgA. In most cases B cells need activation from T helper cells to induce class switching; however, in another pathway, gut flora cause NF- κ B signaling by intestinal epithelial cells which results in further signaling molecules being secreted. These signaling molecules interact with B cells to induce class switching to IgA. IgA is an important type of antibody that is used in mucosal environments like the gut. It has been shown that IgA can help diversify the gut community and helps in getting rid of bacteria that cause inflammatory responses. Ultimately, IgA maintains a healthy environment between the host and gut bacteria. These cytokines and antibodies can have effects outside the gut, in the lungs and other tissues.

The immune system can also be altered due to the gut bacteria's ability to produce metabolites (molecules formed from metabolism) that can effect cells in the immune system. For example short-chain fatty acids (SCFA) can be produced by some gut bacteria through fermentation. SCFAs stimulate a rapid increase in the production of innate immune cells like neutrophils, basophils and eosinophils. These cells are part of the innate immune system that try to limit the spread of infection.

Metabolism

Without gut flora, the human body would be unable to utilize some of the undigested carbohydrates it consumes, because some types of gut flora have enzymes that human cells lack for breaking down certain polysaccharides. Rodents raised in a sterile environment and lacking in gut flora need to eat 30 per cent more calories just to remain the same weight as their normal counterparts. Carbohydrates that humans cannot digest without bacterial help include certain starches, fibre, oligosaccharides, and sugars that the body failed to digest and absorb like lactose in the case of lactose intolerance and sugar alcohols, mucus produced by the gut, and proteins. Bacteria turn carbohydrates they ferment into short-chain fatty acids by a form of fermentation called saccharolytic fermentation. Products include acetic acid, propionic acid and butyric acid. These materials can be used by host cells, providing a major source of energy and nutrients, as well as aiding absorption of essential dietary minerals, such as calcium, magnesium, and iron. Gases and organic acids, such as lactic acid, are also produced by fermentation. Acetic acid is used by muscle, propionic acid facilitates liver production of ATP, and butyric acid provides energy to gut cells. Gut flora also synthesize vitamins like biotin and folate, and facilitate absorption of dietary minerals, including magnesium, calcium, and iron. *Methanobrevibacter smithii* is unique because it is not a species of bacteria, but rather a member of phylum *Archaea*, and is the most abundant methane-producing archaeal species in the human gastrointestinal microbiota.

Host-gut Microbiota-xenobiotic Interaction

Apart from carbohydrates, gut microbiota can also metabolize other xenobiotics such as drugs, phytochemicals, and food toxicants. More than 30 drugs have been shown to be metabolized by gut microbiota. The microbial metabolism of drugs can sometimes inactivate the drug.

Gut-brain Axis

The gut-brain axis is the biochemical signaling that takes place between the gastrointestinal tract and the central nervous system. That term has been expanded to include the role of the gut flora in the interplay; the term “microbiome-gut-brain axis” is sometimes used to describe paradigms explicitly including the gut flora. Broadly defined, the gut-brain axis includes the central nervous system, neuroendocrine and neuroimmune systems including the hypothalamic–pituitary–adrenal axis (HPA axis), sympathetic and parasympathetic arms of the autonomic nervous system including the enteric nervous system, the vagus nerve, and the gut microbiota.

Alterations in Flora Balance

Effects of Antibiotic Use

Altering the numbers of gut bacteria, for example by taking broad-spectrum antibiotics, may affect the host's health and ability to digest food. Antibiotics can cause antibiotic-associated diarrhea (AAD) by irritating the bowel directly, changing the levels of gut flora, or allowing pathogenic bacteria to grow. Another harmful effect of antibiotics is the increase in numbers of antibiotic-resistant bacteria found after their use, which, when they invade the host, cause illnesses that are difficult to treat with antibiotics. Changing the numbers and species of gut flora can reduce the body's ability to ferment carbohydrates and metabolize bile acids and may cause diarrhea. Carbohydrates that are not broken down may absorb too much water and cause runny stools, or lack of SCFAs produced by gut flora could cause diarrhea. A reduction in levels of native bacterial species also disrupts their ability to inhibit the growth of harmful species such as *C. difficile* and *Salmonella kedougou*, and these species can get out of hand, though their overgrowth may be incidental and not be the true cause of diarrhea. Emerging

treatment protocols for *C. difficile* infections involve fecal microbiota transplantation of donor feces.. Initial reports of treatment describe success rates of 90 per cent, with few side effects. Efficacy is speculated to result from restoring bacterial balances of bacteroides and firmicutes classes of bacteria. Gut flora composition also changes in severe illnesses, due not only to antibiotic use but also to such factors as ischemia of the gut, failure to eat, and immune compromise. Negative effects from this have led to interest in selective digestive tract decontamination (SDD), a treatment to kill only pathogenic bacteria and allow the re-establishment of healthy ones. Antibiotics alter the population of the gastrointestinal (GI) tract microbiota, may change the intra-community metabolic interactions, modify caloric intake by using carbohydrates, and globally affects host metabolic, hormonal and immune homeostasis. There is reasonable evidence that taking probiotics containing *Lactobacillus* species may help prevent antibiotic-associated diarrhea and that taking probiotics with *Saccharomyces* (e.g., *Saccharomyces boulardii*) may help to prevent *Clostridium difficile* infection following systemic antibiotic treatment.

Pregnancy

Women's gut microbiota change as pregnancy advances, with the changes similar to those seen in metabolic syndromes such as diabetes. The change in gut flora causes no ill effects. The newborn's gut biota resemble the mother's first-trimester samples. The diversity of the flora decreases from the first to third trimester, as the numbers of certain species go up.

Probiotics, Prebiotics, Synbiotics, and Pharmabiotics

Probiotics are microorganisms that are believed to provide health benefits when consumed. With regard to gut flora, prebiotics are typically non-digestible, fibre compounds that pass undigested through the upper part of the gastrointestinal tract and stimulate the growth or activity of advantageous gut flora by acting as substrate for them. Synbiotics refers to food ingredients or dietary supplements combining probiotics and prebiotics in a form of synergism. The term "pharmabiotics" is used in various ways, to mean: pharmaceutical formulations (standardized manufacturing that can obtain regulatory approval as a drug) of probiotics, prebiotics, or synbiotics; probiotics that have been genetically engineered or otherwise optimized for best performance (shelf life, survival in the digestive tract, etc.); and the natural products of gut flora metabolism (vitamins, etc.).

There is some evidence that treatment with some probiotic strains of bacteria may be effective in irritable bowel syndrome and chronic idiopathic constipation. Those organisms most likely to result in a decrease of symptoms have included:

- *Enterococcus faecium*
- *Lactobacillus plantarum*
- *Lactobacillus rhamnosus*
- *Propionibacterium freudenreichii*
- *Bifidobacterium breve*
- *Lactobacillus reuteri*
- *Lactobacillus salivarius*
- *Bifidobacterium infantis*
- *Streptococcus thermophilus*

Research

Tests for whether non-antibiotic drugs may impact human gut-associated bacteria were performed by *in vitro* analysis on more than 1000 marketed drugs against 40 gut bacterial strains, demonstrating that 24 per cent of the drugs inhibited the growth of at least one of the bacterial strains.

Role in Disease

Bacteria in the digestive tract can contribute to and be affected by the disease in various ways. The presence or overabundance of some kinds of bacteria may contribute to inflammatory disorders such as inflammatory bowel disease. Additionally, metabolites from certain members of the gut flora may influence host signalling pathways, contributing to disorders such as obesity and colon cancer. Alternatively, in the event of a breakdown of the gut epithelium, the intrusion of gut flora components into other host compartments can lead to sepsis.

Ulcers

Helicobacter pylori can cause stomach ulcers by crossing the epithelial lining of the stomach. Here the body produces an immune response. During this response, parietal cells are stimulated and release extra hydrochloric acid (HCl) into the stomach. However, the response does not stimulate the mucus-secreting cells that protect and line the epithelium of the stomach. The extra acid sears holes into the epithelial lining of the stomach, resulting in stomach ulcers.

Inflammatory Bowel Diseases

The two main types of inflammatory bowel diseases, Crohn's disease and ulcerative colitis, are chronic inflammatory disorders of the gut; the causes of these disease are unknown and issues with the gut flora and its relationship with the host have been implicated in these conditions. Additionally, it appears that interactions of gut flora with the gut-brain axis have a role in IBD, with physiological stress mediated through the hypothalamic–pituitary–adrenal axis driving changes to intestinal epithelium and the gut flora in turn releasing factors and metabolites that trigger signaling in the enteric nervous system and the vagus nerve. The diversity of gut flora appears to be significantly diminished in people with inflammatory bowel diseases compared to healthy people; additionally, in people with ulcerative colitis, Proteobacteria and Actinobacteria appear to dominate; in people with Crohn's, *Enterococcus faecium* and several Proteobacteria appear to be over-represented. There is reasonable evidence that correcting gut flora imbalances by taking probiotics with *Lactobacilli* and *Bifidobacteria* can reduce visceral pain and gut inflammation in IBD.

Irritable Bowel Syndrome

Irritable bowel syndrome is a result of stress and chronic activation of the HPA axis; its symptoms include abdominal pain, changes in bowel movements, and an increase in proinflammatory cytokines. Overall, studies have found that the luminal and mucosal microbiota are changed in irritable bowel syndrome individuals, and these changes can relate to the type of irritation such as diarrhea or constipation. Also, there is a decrease in the diversity of the microbiome with low levels of fecal *Lactobacilli* and *Bifidobacteria*, high levels of facultative anaerobic bacteria such as *Escherichia coli*, and increased ratios of Firmicutes:Bacteroidetes.

Other Inflammatory or Autoimmune Conditions

Allergy, asthma, and diabetes mellitus are autoimmune and inflammatory disorders of unknown cause, but have been linked to imbalances in the gut flora and its relationship with the host. With asthma, two hypotheses have been posed to explain its rising prevalence in the developed world. The hygiene hypothesis posits that children in the developed world are not exposed to enough microbes and thus may contain lower prevalence of specific bacterial taxa that play protective roles. The second hypothesis focuses on the Western pattern diet, which lacks whole grains and fibre and has an overabundance of simple sugars. Both hypotheses converge on the role of short-chain fatty acids (SCFAs) in immunomodulation. These bacterial fermentation metabolites are

involved in immune signalling that prevents the triggering of asthma and lower SCFA levels are associated with the disease. Lacking protective genera such as *Lachnospira*, *Veillonella*, *Rothia* and *Faecalibacterium* has been linked to reduced SCFA levels. Further, SCFAs are the product of bacterial fermentation of fibre, which is low in the Western pattern diet. SCFAs offer a link between gut flora and immune disorders, and as of 2016, this was an active area of research. Similar hypotheses have also been posited for the rise of food and other allergies. The connection between the gut microbiota and diabetes mellitus type 1 has also been linked to anti-inflammatory SCFAs, such as butyrate, regulating the associated low-grade inflammatory state by controlling gut permeability. Additionally, butyrate has also been shown to decrease insulin resistance, suggesting gut communities low in butyrate-producing microbes may increase chances of acquiring diabetes mellitus type 2. As of 2016 it was not clear if changes to the gut flora cause these auto-immune and inflammatory disorders or are a product of or adaptation to them.

Obesity and Metabolic Syndrome

The gut flora has also been implicated in obesity and metabolic syndrome due to the key role it plays in the digestive process; the Western pattern diet appears to drive and maintain changes in the gut flora that in turn change how much energy is derived from food and how that energy is used. One aspect of a healthy diet that is often lacking in the Western-pattern diet is fibre and other complex carbohydrates that a healthy gut flora require flourishing; changes to gut flora in response to a Western-pattern diet appear to increase the amount of energy generated by the gut flora which may contribute to obesity and metabolic syndrome. There is also evidence that microbiota influence eating behaviours based on the preferences of the microbiota, which can lead to the host consuming more food eventually resulting in obesity. It has generally been observed that with higher gut microbiome diversity, the microbiota will spend energy and resources on competing with other microbiota and less on manipulating the host. The opposite is seen with lower gut microbiome diversity, and these microbiotas may work together to create host food cravings. Additionally, the liver plays a dominant role in blood glucose homeostasis by maintaining a balance between the uptake and storage of glucose through the metabolic pathways of glycogenesis and gluconeogenesis. In recent studies, it is illustrated that intestinal lipids regulate glucose homeostasis involving a gut-brain-liver axis. The direct administration of lipids into the upper intestine increases the long chain fatty acyl-coenzyme A (LCFA-CoA) levels in the upper intestines and suppresses glucose production even under subdiaphragmatic vagotomy or gut vagal deafferentation. This interrupts the neural connection between the brain and the gut and blocks the upper intestinal lipids' ability to inhibit glucose production. The gut-brain-liver axis and gut microbiota composition can regulate the glucose homeostasis in the liver and provide potential therapeutic methods to treat obesity and diabetes. Just as gut flora can function in a feedback loop that can drive the development of obesity, there is evidence that restricting intake of calories (*i.e.*, dieting) can drive changes to the composition of the gut flora.

Liver Disease

As the liver is fed directly by the portal vein, whatever crosses the intestinal epithelium and the intestinal mucosal barrier enters the liver, as do cytokines generated there. Dysbiosis in the gut flora has been linked with the development of cirrhosis and non-alcoholic fatty liver disease.

Systemic Infections

Normally-commensal bacteria can be very harmful to the host if they get outside of the intestinal tract. Translocation, which occurs when bacteria leave the gut through its mucosal lining, the border between the lumen of the gut and the inside of the body, can occur in a number of different diseases, and can be caused by

too much growth of bacteria in the small intestine, reduced immunity of the host, or increased gut lining permeability. If the gut is perforated, bacteria can invade the body, causing a potentially fatal infection. Aerobic bacteria can make an infection worse by using up all available oxygen and creating an environment favourable to anaerobes.

Cancer

Some genera of bacteria, such as *Bacteroides* and *Clostridium*, have been associated with an increase in tumor growth rate, while other genera, such as *Lactobacillus* and *Bifidobacteria*, are known to prevent tumor formation. As of December 2017 there was preliminary and indirect evidence that gut microbiota might mediate response to PD-1 inhibitors; the possible mechanism was unknown.

Neuropsychiatric

Interest in the relationship between gut flora and neuropsychiatric issues was sparked by a 2004 study showing that germ-free mice showed an exaggerated HPA axis response to stress compared to non-GF laboratory mice. As of January 2016, most of the work that has been done on the role of gut flora in the gut-brain axis had been conducted in animals, or characterizing the various neuroactive compounds that gut flora can produce, and studies with humans measuring differences between people with various psychiatric and neurological differences, or changes to gut flora in response to stress, or measuring effects of various probiotics (dubbed “psychobiotics in this context), had generally been small and could not be generalized; whether changes to gut flora are a result of disease, a cause of disease, or both in any number of possible feedback loops in the gut-brain axis, remained unclear. A systematic review from 2016 examined the preclinical and small human trials that have been conducted with certain commercially available strains of probiotic bacteria and found that among those tested, *Bifidobacterium* and *Lactobacillus* genera (*B. longum*, *B. breve*, *B. infantis*, *L. helveticus*, *L. rhamnosus*, *L. plantarum*, and *L. casei*), had the most potential to be useful for certain central nervous system disorders.

Other Animals

Aside from mammals, some insects also possess complex and diverse gut microbiota that play key nutritional roles. Microbial communities associated termites can constitute a majority of the weight of the individuals and perform important roles in the digestion of lignocellulose and nitrogen fixation. These communities are host-specific, and closely related insect species share comparable similarities in gut microbiota composition. In cockroaches, gut microbiota have been shown to assemble in a deterministic fashion, irrespective of the inoculum; the reason for this host-specific assembly remains unclear. Bacterial communities associated with insects like termites and cockroaches are determined by a combination of forces, primarily diet, but there is some indication that host phylogeny may also be playing a role in the selection of lineages. For more than 51 years it has been known that the administration of low doses of antibacterial agents promotes the growth of farm animals to increase weight gain.

In a study performed on mice by Ilseung Cho, the ratio of *Firmicutes* and *Lachnospiraceae* was significantly elevated in animals treated with subtherapeutic doses of different antibiotics. By analyzing the caloric content of faeces and the concentration of small chain fatty acids (SCFAs) in the GI tract, they concluded that the changes in the composition of microbiota lead to an increased capacity to extract calories from otherwise indigestible constituents, and to an increased production of SCFAs. These findings provide evidence that antibiotics perturb not only the composition of the GI microbiome but also its metabolic capabilities, specifically with respect to SCFAs.

VAGINA

Vaginal flora or vaginal microbiota are the microorganisms that colonize the vagina. They were discovered by the German gynecologist Albert Döderlein in 1892, and are part of the overall human flora. The amount and type of bacteria present have significant implications for a woman's overall health. The primary colonizing bacteria of a healthy individual are of the genus *Lactobacillus*, such as *L. crispatus*, and the lactic acid they produce is thought to protect against infection by pathogenic species.

Lactobacilli

The primary colonizing bacteria of a healthy individual are of the genus *Lactobacillus*. Since the first description of lactobacilli by Döderlein, lactobacilli have been generally considered as the gatekeepers of the vaginal ecosystem. Lactobacilli have been shown to inhibit in vitro growth of pathogenic microorganisms, e.g. *Bacteroides fragilis*, *Escherichia coli*, *Gardnerella vaginalis*, *Mobiluncus* spp., *Neisseria gonorrhoeae*, *Peptostreptococcus anaerobius*, *P. bivia* and *Staphylococcus aureus*. It is generally accepted that this is achieved mainly through the action of lactic acid. Moreover, lactobacilli normally help to prevent long-term colonization of the vagina by adhering to vaginal epithelial cells. This usually reduces pathogens from infecting to the vaginal epithelium. Next to lactic acid production and competition for adherence, other antagonistic mechanisms include hydrogen peroxide (a broad-spectrum antimicrobial) and bacteriocins (target-specific antimicrobials) production.

The Production of Lactic Acid

Low pH is generally accepted to be the main mechanism controlling the composition of the vaginal microflora. Although the lactic acid produced by lactobacilli contributes to the vaginal acidity, it is still not proven to be the primary source of low vaginal pH, but the fact remains that most lactobacilli thrive best at a pH <4.5.

Hydrogen Peroxide

Production of hydrogen peroxide (H_2O_2) is a well-known mechanism for bacterial antagonism, inhibiting growth of microorganisms via direct interaction or via human myeloperoxidase. Hydrogen peroxide-producing lactobacilli have been shown to inactivate HIV-1, herpes simplex virus type 2 (HSV-2), *Trichomonas vaginalis*, *G. vaginalis*, *P. bivia* and *E. coli*. O'Hanlon and Baeten found that 96 per cent of *Lactobacillus* species from a healthy vaginal ecosystem produced H_2O_2 (*L. jensenii* and *L. vaginalis* produce the highest levels of H_2O_2), whereas only 6 per cent of the lactobacilli recovered from women with BV produced H_2O_2 . In agreement with this, *L. iners*, most frequently associated with disturbed vaginal microflora, is a poor producer of H_2O_2 . Vaginal colonization by H_2O_2 -producing lactobacilli has been associated with a decrease in the occurrence of bacterial vaginosis (BV). However, more recently O'Hanlon et al. demonstrated that cervicovaginal fluid and semen have a significant H_2O_2 -blocking activity and they later demonstrated that physiological concentrations of H_2O_2 below 100 μM fail to inactivate any of the 17 tested BV-associated bacteria, e.g. *A. vaginae*, *G. vaginalis*, *Mobiluncus* spp., *P. bivia*, *Prevotella corporis*, *Mycoplasma hominis*, even in the presence of human myeloperoxidase, known to increase the microbicidal activity of H_2O_2 . Only supraphysiologic concentrations of exogenous H_2O_2 (0.34 per cent w/v, 100 mM) were sufficient to inactivate BV-associated bacteria at which concentration it more potently inactivated vaginal lactobacilli (*L. crispatus*, *L. gasseri*, *L. iners* and *L. jensenii*). A concentration of 100 mM H_2O_2 is approximately 50-fold higher than lactobacilli are capable of producing even under optimal aerobic, low-antioxidant conditions, and approximately 5,000-fold higher than the estimated H_2O_2 concentration in vivo. Even more remarkable, the addition of only 1 per cent vaginal fluid blocked the microbicidal activity of 1 M H_2O_2 . Possible explanations may be that cervicovaginal fluid and semen contain

proteins, glycoproteins, polysaccharides, lipids, and other molecules with the potential to react with and inactivate H_2O_2 . In addition, the vagina is hypoxic most of the time, whereas lactobacilli require oxygen to produce hydrogen peroxide. It is also remarkable that catalase, which provides bacteria protection against toxic H_2O_2 , is absent in lactobacilli, and as such they would be unprotected against their own H_2O_2 production. In contrast, under optimal anaerobic growth conditions, physiological concentrations of lactic acid inactivated the BV-associated pathogens without affecting the vaginal lactobacilli.

In summary, although the hydrogen peroxide production of lactobacilli has been considered as an important antimicrobial component, contributing to the colonization resistance provided by lactobacilli, and although there seems to be a link between H_2O_2 -producing lactobacilli and normal vaginal microflora, recent data do not support this role for H_2O_2 .

Bacteriocins

Vaginal lactobacilli produce antimicrobial peptides, *i.e.* bacteriocins such as lactocin 160 and crispsin, with inhibitory activity ranging from narrow (closely related *Lactobacillus* species) to broad (diverse groups of bacteria, including *G. vaginalis* and *P. bivia*), and bacteriocin-like substances, with a broader spectrum of activity than bacteriocins (*e.g.* a heat-resistant peptide produced by *Lactobacillus salivarius* subsp. *salivarius* CRL 1328). Several studies have indicated that the activity of bacteriocins is favored by low pH. The inhibitory substances produced by vaginal *Lactobacillus* is a primary factor in protecting the vaginal microbiota, with organic acids, bacteriocins, and hydrogen peroxide. These act synergistically against infection by pathogens. Not all *Lactobacillus* spp. and not all strains within one *Lactobacillus* species exhibit all 3 mechanisms. *Lactobacillus* species differ in premenopausal women, *i.e.* *L. crispatus*, *L. jensenii*, *L. iners*, *L. gasseri* (and possibly *L. vaginalis*), as assessed through cultivation-dependent and cultivation-independent techniques. Vaginal lactobacilli have been shown to display a pronounced vaginotropism, and their pili act as ligands for attachment to receptors of vaginal epithelial cells. The limited number of *Lactobacillus* spp. found in the human vagina is remarkable, which leads to the possibility that there are host factors that select for specific organisms, that these species have unusual characteristics that allow them to successfully colonize the vagina, or both. However, the vaginotropism, does not only apply to this selected group of lactobacilli that stand for a healthy vagina, but also for the bacterial species associated with BV. The microbiota detected in the human genital and gut ecoiniche do not appear to grow outside their host and probably are likely to rely on the close contact between parents and their children for transmission, *e.g.* mother to neonate transmission of genital microflora, most probably also with gut microflora homogeneously distributed over the baby's body including skin, the oral cavity, nasopharynx, and feces.

Other Microbiota

Healthy, normal vaginal microbiota that is dominated by lactobacilli may differ among some ethnic groups. Non-pathogenic vaginal species are part of the normal microbiota of some women. Several studies have demonstrated that a significant proportion (7–33 per cent) of healthy asymptomatic women (especially black and Hispanic women) lack appreciable numbers of *Lactobacillus* species in the vagina, and instead have a vaginal microbiota that consist of other lactic acid-producing bacteria, *i.e.* species from the genera *Atopobium*, *Leptotrichia*, *Leuconostoc*, *Megasphaera*, *Pediococcus*, *Streptococcus* and *Weissella*. All ethnic populations have vaginal microflora communities containing lactic acid producing bacteria. This implies that not all communities may be equally resilient, so that if the resilience of a vaginal community is low then transitory changes in the structure of these communities may occur more readily in response to disturbances of various kinds, including menses, sexual intercourse, douching and contraceptive practices. These differences in the structure and composition of microbial communities may underlie well-known differences in the susceptibility

of women in these racial groups to BV and various vaginal infections. Though vaginal microflora may be disturbed, and lactic acid is produced by other species, vaginal pH and acid production establish normal vaginal microbiota. The pH further decreases during pregnancy.

Other Vaginal Bacterial Species

Other bacterial species are frequently found in the vagina, such as the Gram positive cocci: *Atopobium vaginae*, *Peptostreptococcus* spp., *Staphylococcus* spp., *Streptococcus* spp., and *Bacteroides* spp., *Fusobacterium* spp., *Gardnerella vaginalis*, *Mobiluncus*, *Prevotella* spp., and Gram-negative enteric organisms, such as *Escherichia coli*. Mycoplasma and Ureaplasma are frequently found in the vagina. Some of the obligate and facultative anaerobic bacteria are associated with BV.

Pregnancy

The effect of tampon use on vaginal flora is debated, but application of tampons appears not to significantly modify the balance of bacterial presence. Pregnancy alters the microbiota with a reduction in species/genus diversity.

Disease Prevention

A healthy vaginal microbiome aids in the prevention of bacterial vaginosis, yeast infections and other possible problems by maintaining an acidic pH (< 4.5) that is unfavourable for the growth of common pathogens, such as *Gardnerella vaginalis*. The lactobacilli present in a healthy vaginal microbiome also occupy the ecological niche that would otherwise be available for exploitation by pathogenic organisms. However, harmful bacteria or an imbalance in bacteria can lead to infection. Bacterial vaginosis is associated with the presence of *Gardnerella vaginalis* and *Peptostreptococcus anaerobius* and a decrease in the number of *Lactobacillus* species that comprise the healthy vaginal microbiota.

Research

Investigations have found that the presence of lacto-bacillus dominated bacteria in the vagina is associated with a lower incidence of sexually transmitted infections.

PLACENTA

The placental microbiome is the nonpathogenic, commensal bacteria claimed to be present in a healthy human placenta and is distinct from bacteria that cause infection and preterm birth in chorioamnionitis. Until recently, the healthy placenta was considered to be a sterile organ but now genera and species have been identified that reside in the basal layer. It should be stressed that the evidence for a placental microbiome is controversial. Most studies supporting the existence of a placental microbiome lack the appropriate experimental controls, and it has been found that contamination is most likely responsible for reports of a placental microbiome. The placental microbiome more closely resembles that of the oral microbiome than either the vaginal or rectal microbiome. Changes in the microbiome can result in disease.

Bacterial Species and Genera

Culturable and non-culturable bacterial species in the placenta obtained following normal term pregnancy have been identified.

Binomial name	Commensal	Transient	Potential pathogen	References
<i>Prevotella tannerae</i>	x		x	
<i>Firmicutes spp</i>		x	x	
<i>Tenericutes spp</i>	x		x	
<i>Fusobacterium nucleatum</i>	x		x	
<i>Prevotella tanerae</i>	x			
<i>Bacteroides spp</i>	x		x	
<i>Fusobacterium spp</i>	x		x	
<i>Streptomyces avermitilis</i>	x			
<i>Neisseria polysaccharea</i>	x			
<i>Neisseria lactamica</i>	x			
<i>Proteobacteria spp</i>	x			
<i>Bacteroidetes spp</i>	x			
<i>Escherichia coli</i>	x		x	
<i>Escherichia spp</i>	x		x	
<i>Actinobacteria spp</i>	x		x	
<i>Cyanobacteria spp</i>	x			
<i>Chloroflexi spp</i>	x			
<i>Aquificae spp</i>	x			
<i>Verrucomicrobia spp</i>	x			
<i>Vibrio spp</i>	x			
<i>Burkholderia spp</i>	x			
<i>Beijerinckia spp</i>	x			

In a healthy placental microbiome, the diversity of the species and genera is extensive. A change in the composition of the microbiota in the placenta is associated with excess gestational weight gain, and pre-term birth. The placental microbiota varies between low birth weight infants and those infants with normal birth weights.

While bacteria are often found in the amniotic fluid of failed pregnancies, they are also found in particulate matter that is found in about 1 per cent of healthy pregnancies. In non-human animals, part of the microbiome is passed onto offspring even before the offspring are born. Bacteriologists assume that the same probably holds true for humans.

Research

Future research may find that the microbiota of the female reproductive tract may be related to pregnancy, conception, and birth. Animal studies have been used to investigate the relationship between oral microbiota and the placental microbiota. Mice inoculated with species of oral bacteria demonstrated placental colonization soon afterwards.

History

Investigations into reproductive-associated microbiomes began around 1885 by Theodor Escherich. He wrote that meconium from the newborn was free of bacteria. This was interpreted as the uterine environment was sterile. Other investigations used sterile diapers for meconium collection. No bacteria were able to be cultured from the samples.

Bacteria were detected and were directly proportional to the time between birth and the passage of meconium. A 1927 study demonstrated the presence of bacteria in the amniotic fluid of those that were in labour for longer than six hours.

UTERUS

The uterine microbiome is the commensal, nonpathogenic, bacteria, viruses, yeasts/fungi present in a healthy uterus, amniotic fluid and endometrium and the specific environment which they inhabit. It has been only recently confirmed that the uterus and its tissues are not sterile. Due to improved 16S rRNA gene sequencing techniques, detection of bacteria that are present in low numbers is possible. Using this procedure that allows the detection of bacteria that cannot be cultured outside the body, studies of microbiota present in the uterus are expected to increase. Characteristics. Bacteria, viruses and one genus of yeasts are a normal part of the uterus before and during pregnancy. The uterus has been found to possess its own characteristic microbiome that differs significantly from the vaginal microbiome. Despite its close spatial connection with the vagina, the microbiome of the uterus more closely resembles the commensal bacteria found in the oral cavity. In addition, the immune system is able to differentiate between those bacteria normally found in the uterus and those that are pathogenic. Hormonal changes have an effect on the microbiota of the uterus.

Taxa

Commensals

The organisms listed below have been identified as commensals in the healthy uterus. Some also have the potential for growing to the point of causing disease:

Organism	Commensal	Transient	Potential pathogen	References
<i>Escherichia coli</i>	x		x	
<i>Escherichia</i> spp.		x	x	
<i>Ureaplasma parvum</i>	x		x	
<i>Fusobacterium nucleatum</i>	x			
<i>Prevotella tannerae</i>	x			
<i>Bacteroides</i> spp.	x			
<i>Streptomyces avermitilis</i>	x			
<i>Mycoplasma</i> spp.	x		x	
<i>Neisseria lactamica</i>	x			
<i>Neisseria polysaccharea</i>	x			
Epstein-Barr virus	x		x	
Respiratory-Syncytial virus	x		x	
Adenovirus	x		x	
<i>Candida</i> spp.	x		x	

Pathogens

Other taxa can be present, without causing disease or an immune response. Their presence is associated with negative birth outcomes.

Clinical Significance

Prophylactic antibiotics have been injected into the uterus to treat infertility. This has been done before the transfer of embryos with the intent to improve implantation rates. No association exists between successful implantation and antibiotic treatment. Infertility treatments often progress to the point where a microbiological analysis of the uterine microbiota is performed. Preterm birth is associated with certain species of bacteria that are not normally part of the healthy uterine microbiome.

Pathogenic Organism	Increased risk of	References
Ureaplasma urealyticum	Premature, preterm rupture of membranes Preterm labor cesarean section Placental inflammation Congenital pneumonia bacteremia meningitis fetal lung injury death of infant	
Haemophilus influenzae	Premature, preterm rupture of membranes preterm labor preterm birth	
Ureaplasma parvum		
Fusobacterium nucleatum		
Prevotella tannerae		
Bacteroides spp.		
Streptomyces avermitilis		
Mycoplasma hominis	Congenital pneumonia bacteremia meningitis <pelvic inflammatory disease postpartum or postabortal fever	
Neisseria lactamica		
Neisseria polysaccharea		
Epstein-Barr virus		
Respiratory-Syncytial virus		
Adenovirus		
Candida spp.		

Immune Response

The immune response becomes more pronounced when bacteria are found that are not commensal.

History

Investigations into reproductive-associated microbiomes began around 1885 by Theodor Escherich. He wrote that meconium from the newborn was free of bacteria. There was a general consensus at the time and even recently that the uterus was sterile and this was referred to as the *sterile womb paradigm*. Other investigations used sterile diapers for meconium collection. No bacteria were able to be cultured from the samples. Other studies showed that bacteria were detected and were directly proportional to the time between birth and the passage of meconium.

Research

Investigations into the role of the uterine microbiome in the development of the infant microbiome are ongoing.

ORAL CAVITY

Oral microbiology is the study of the microorganisms (microbiota) of the oral cavity and their interactions between oral microorganisms or with the host. The environment present in the human mouth allows the growth of characteristic microorganisms found there. It provides a source of water and nutrients, as well as a moderate

temperature. Resident microbes of the mouth adhere to the teeth and gums to resist mechanical flushing from the mouth to stomach where acid-sensitive microbes are destroyed by hydrochloric acid. Anaerobic bacteria in the oral cavity include: *Actinomyces*, *Arachnia*, *Bacteroides*, *Bifidobacterium*, *Eubacterium*, *Fusobacterium*, *Lactobacillus*, *Leptotrichia*, *Peptococcus*, *Peptostreptococcus*, *Propionibacterium*, *Selenomonas*, *Treponema*, and *Veillonella*. Genera of fungi that are frequently found in the mouth include *Candida*, *Cladosporium*, *Aspergillus*, *Fusarium*, *Glomus*, *Alternaria*, *Penicillium*, and *Cryptococcus*, among others. Bacteria accumulate on both the hard and soft oral tissues in biofilms. Bacterial adhesion is particularly important for oral bacteria. Oral bacteria have evolved mechanisms to sense their environment and evade or modify the host. Bacteria occupy the ecological niche provided by both the tooth surface and gingival epithelium. However, a highly efficient innate host defence system constantly monitors the bacterial colonization and prevents bacterial invasion of local tissues. A dynamic equilibrium exists between dental plaque bacteria and the innate host defence system. Of particular interest is the role of oral microorganisms in the two major dental diseases: dental caries and periodontal disease. Additionally, research has correlated poor oral health and the resulting ability of the oral microbiota to invade the body to affect cardiac health as well as cognitive function.

Oral Bacteria

The environment present in the human mouth allows the growth of characteristic microorganisms found there. It provides a source of water and nutrients, as well as a moderate temperature. Resident microbes of the mouth adhere to the teeth and gums to resist mechanical flushing from the mouth to stomach where acid-sensitive microbes are destroyed by hydrochloric acid. Anaerobic bacteria in the oral cavity include: *Actinomyces*, *Arachnia*, *Bacteroides*, *Bifidobacterium*, *Eubacterium*, *Fusobacterium*, *Lactobacillus*, *Leptotrichia*, *Peptococcus*, *Peptostreptococcus*, *Propionibacterium*, *Selenomonas*, *Treponema*, and *Veillonella*. The oral cavity of the newborn baby does not contain bacteria but rapidly becomes colonized with bacteria such as *Streptococcus salivarius*. With the appearance of the teeth during the first year colonization by *Streptococcus mutans* and *Streptococcus sanguinis* occurs as these organisms colonise the dental surface and gingiva. Other strains of streptococci adhere strongly to the gums and cheeks but not to the teeth. The gingival crevice area (supporting structures of the teeth) provides a habitat for a variety of anaerobic species. *Bacteroides* and spirochetes colonize the mouth around puberty. Of particular interest is the role of oral microorganisms in the two major dental diseases: dental caries and periodontal disease.

Fusospirochetes

Spirochetes and fusiform bacilli live as normal flora in the mouth, but in the case of bleeding in the oral cavity, the bacteria can cause infection and diseases to oral cavity:

- Acute necrotizing ulcerative gingivitis (ANUG)
- Vincent angina with a membrane covering the throat area

Veillonella

Veillonella are gram-negative anaerobic cocci. It is thought that this species thrives in the acidic environment of caries and is thought to slow the development of dental caries. It converts the acidic products of other species to less acidic products.

Actinobacillus Actinomycetemcomitans

Actinobacillus actinomycetemcomitans is considered an oral pathogen due to its virulence factors, its association with localized aggressive periodontitis in young adolescents, and studies indicating that it can cause bone loss.

Lactobacillus

Some *Lactobacillus* species have been associated with dental cavities although these bacteria are normally symbiotic in humans and are found in the gut flora.

Dental Plaque

Dental plaque is the material that adheres to the teeth and consists of bacterial cells (mainly *S. mutans* and *S. sanguis*), salivary polymers and bacterial extracellular products. Plaque is a biofilm on the surfaces of the teeth. This accumulation of microorganisms subject the teeth and gingival tissues to high concentrations of bacterial metabolites which results in dental disease. If not taken care of, via brushing or flossing, the plaque can turn into tartar (its hardened form) and lead to gingivitis or periodontal disease.

Cell–cell Communication

Most of the bacterial species found in the mouth belong to microbial communities, called biofilms, a feature of which is inter-bacterial communication. Cell–cell contact is mediated by specific protein adhesins and often, as in the case of inter-species aggregation, by complementary polysaccharide receptors. Another method of communication involves cell–cell signalling molecules, which are of two classes: those used for intra-species and those used for inter-species signalling. An example of intra-species communication is quorum sensing. Oral bacteria have been shown to produce small peptides, such as competence stimulating peptides, which can help promote single-species biofilm formation. A common form of inter-species signalling is mediated by 4, 5-dihydroxy-2, 3-pentanedione (DPD), also known as autoinducer-2 (AI-2).

Vaccination Against Oral Infections

In the case of dental caries, proteins involved in colonization of teeth by *Streptococcus mutans* can produce antibodies that inhibit the cariogenic process.

Other Microbiota

Bacteria species typically associated with oral flora have been found to be present in women with bacterial vaginosis. Genera of fungi that are frequently found in the mouth include *Candida*, *Cladosporium*, *Aspergillus*, *Fusarium*, *Glomus*, *Alternaria*, *Penicillium*, and *Cryptococcus*, among others.

Role in Non-oral Disease

Additionally, research has correlated poor oral health and the resulting ability of the oral microbiota to invade the body to affect cardiac health as well as cognitive function.

LUNG

The lung microbiota, is the pulmonary microbial community consisting of a complex variety of microorganisms found in the lower respiratory tract particularly on the mucous layer and the epithelial surfaces. These microorganisms include bacteria, fungi, viruses and bacteriophages. The bacterial part of the microbiota has been more closely studied. It consists of a core of nine genera: *Prevotella*, *Sphingomonas*, *Pseudomonas*, *Acinetobacter*, *Fusobacterium*, *Megasphaera*, *Veillonella*, *Staphylococcus*, and *Streptococcus*. They are aerobes as well as anaerobes and aerotolerant bacteria. The microbial communities are highly variable in particular individuals and compose of about 140 distinct families. The bronchial tree for instance contains a mean of 2000

bacterial genomes per cm surface. The harmful or potentially harmful bacteria are also detected routinely in respiratory specimens. The most significant are *Moraxella catarrhalis*, *Haemophilus influenzae*, and *Streptococcus pneumoniae*. They are known to cause respiratory disorders under particular conditions namely if the human immune system is impaired. The mechanism by which they persist in the lower airways in healthy individuals is unknown. Fungal genera that are commonly found in the lung microbiota include *Candida*, *Malassezia*, *Neosartorya*, *Saccharomyces*, and *Aspergillus*, among others.

Role of the Epithelial Barrier

The airway epithelium together with alveolar macrophages and dendritic cells play a major role in the initial recognition of bacterial products getting into the lower airways with the air. Since some of these products are potent proinflammatory stimuli it is extremely important for the immune system to distinguish between pathogens and non-pathogenic commensals. This prevents the development of constant inflammation and forms tolerance against harmless microbiota. This process becomes much more intriguing when taking into account that commensals often share their surface molecules with pathogens. Epithelial cells are equipped with very sensitive recognition tools - toll like receptors (TLRs), nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) and retinoic acid-inducible gene (RIG)-I-like receptors (RLRs) which recognize a broad variety of microbial structural components. After recognition of pathogenic bacteria proinflammatory pathways are activated and cellular components of the adaptive and innate immunity are recruited to the infection site. One key regulator in this process is the NF- κ B which translocates from the cytoplasm into the nucleus and activates pro-inflammatory genes in epithelial cells and macrophages. The DNA-binding protein complex recognizes a discrete nucleotide sequence (5'-GGG ACT TTC T-3') in the upstream region of a variety of response genes. The activation of NF- κ B by a number of stimuli: bacterial cell walls or inflammatory cytokines results in its translocation to the nucleus. In contrast, harmless bacteria do not cause the translocation of NF- κ B into the nucleus thus preventing the inflammation although they can express the same microbe-associated molecular patterns (MAMPs). One possible mechanism explaining this effect was suggested by Neish showing that non-pathogenic *S. typhimurium* PhoPc and *S. pullorum* are able to prohibit the ubiquitination of NF- κ B inhibitor molecule nuclear factor of NF- κ B light polypeptide gene enhancer in B-cells inhibitor alpha (I κ B- κ). Another explanation of commensal tolerance of the epithelium refers to the post-translational modification of a protein by the covalent attachment of one or more ubiquitin (Ub) monomers. The inhibition of ubiquitination leads to reduction of inflammation, because only polyubiquitinated (I κ B- κ is targeted for degradation by the 26 S proteasome, allowing NF- κ B translocation to the nucleus and activation the transcription of effector genes (for example IL-8).

Probiotic bacteria such as *Lactobacilli* are able to modulate the activity of the Ub-proteasome system via inducing reactive oxygen species (ROS) production in epithelial cells. In mammalian cells, ROS have been shown to serve as critical second messengers in multiple signal transduction pathways in response to proinflammatory cytokines. Bacterially induced ROS causes oxidative inactivation of the catalytic cysteine residue of Ub 12 resulting in incomplete but transient loss of cullin-1 neddylation and consequent effects on NF- κ B and β -catenin signaling. Another commensal species, *B. thetaiotaomicron*, attenuates pro-inflammatory cytokine expression by promoting nuclear export of NF- κ B subunit RelA, through a peroxisome proliferator activated receptor γ (PPAR- γ)-dependent pathway. PPAR- γ target transcriptionally active Rel A and induce early nuclear clearance limiting the duration of NF- κ B action. The balance between pathogens and commensals is extremely important in the maintenance of homeostasis in the respiratory tract.

Physiology

The airways are continually exposed to a multitude of microorganisms, some of which are able to persist and even colonize respiratory tract. This is possible due to the presence of nutrients, oxygen, and optimal growth

temperature. There are several host-derived nutrient sources for microbial residents: secretions from airway epithelial cells (especially goblet cells), secretions from submucosal glands and transudate from plasma. Moreover, the pool of available nutrients is increased by the activities of some members of the microbiota. Macromolecular components of respiratory secretions (proteins, glycoproteins, lipids, nucleic acids) are converted to nutrients (*e.g.* carbohydrates, amino acids). Thus, the metabolic activity of present bacteria allow for the colonization of new species. The commensal bacteria are nonpathogenic and defend our airways against the pathogens. There are several possible mechanisms. Commensals are the native competitors of pathogenic bacteria, because they tend to occupy the same ecological niche inside the human body. Secondly, they are able to produce antibacterial substances called bacteriocins which inhibit the growth of pathogens. Genera *Bacillus*, *Lactobacillus*, *Lactococcus*, *Staphylococcus*, *Streptococcus*, and *Streptomyces* are the main producers of bacteriocins in respiratory tract. Moreover, commensals are known to induce Th1 response and anti-inflammatory interleukin (IL)-10, antimicrobial peptides, FOXP3, secretory immunoglobulin A (sIgA) production.

Clinical Significance

Changes in microbial community composition seem to play a role in progression of such pulmonary disorders as chronic obstructive pulmonary disease (COPD), asthma, and cystic fibrosis. In humans, *S. aureus* is part of the normal microbiota present in the upper respiratory tract, and on skin and in the gut mucosa. *S. aureus*, along with similar species that can colonize and act symbiotically but can cause disease if they begin to take over the tissues they have colonized or invade other tissues, have been called “pathobionts”. MRSA can similarly colonize people without making them sick. The presence of such genera as *Mycoplasma*, *Pseudomonas*, and *Staphylococcus* is correlated with stable COPD state. On the other hand, *Prevotella*, *Mesorhizobium*, *Microbacterium*, *Micrococcus*, *Veillonella*, *Rhizobium*, *Stenotrophomonas*, and *Lactococcus* present mostly in healthy individual cohort. The relative abundance of *Proteobacteria* is increased in asthmatic children. *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Burkholderia cepacia* are found most often in cystic fibrosis patients. High-throughput sequencing and the whole genome sequencing approaches will provide the further information about the complexity and physiological implication of commensal bacteria in the lower respiratory tract.

BILIARY TRACT

Traditionally the biliary tract has been considered to be normally sterile, and the presence of microorganisms in bile is a marker of pathological process. This assumption was confirmed by failure in allocation of bacterial strains from the normal bile duct. Papers began emerging in 2013 showing that the normal biliary microbiota is a separate functional layer which protects a biliary tract from colonization by exogenous microorganisms.

DISEASE AND DEATH

A symbiotic relationship between the gut microbiota and different bacteria may influence an individual's immune response.

CANCER

Although cancer is generally a disease of host genetics and environmental factors, microorganisms are implicated in some 20 per cent of human cancers. Particularly for potential factors in colon cancer, bacterial density is one million times higher than in the small intestine, and approximately 12-fold more cancers occur in the colon compared to the small intestine, possibly establishing a pathogenic role for microbiota in colon and rectal cancers. Microbial density may be used as a prognostic tool in assessment of colorectal cancers.

The microbiota may affect carcinogenesis in three broad ways:

- Altering the balance of tumor cell proliferation and death,
- Regulating immune system function, and
- Influencing metabolism of host-produced factors, foods and pharmaceuticals. Tumors arising at boundary surfaces, such as the skin, oropharynx and respiratory, digestive and urogenital tracts, harbour a microbiota. Substantial microbe presence at a tumor site does not establish association or causal links. Instead, microbes may find tumor oxygen tension or nutrient profile supportive. Decreased populations of specific microbes or induced oxidative stress may also increase risks. Of the around 10 microbes on earth, ten are designated by the International Agency for Research on Cancer as human carcinogens. Microbes may secrete proteins or other factors directly drive cell proliferation in the host, or may up- or down-regulate the host immune system including driving acute or chronic inflammation in ways that contribute to carcinogenesis.

Concerning the relationship of immune function and development of inflammation, mucosal surface barriers are subject to environmental risks and must rapidly repair to maintain homeostasis. Compromised host or microbiota resiliency also reduce resistance to malignancy, possibly inducing inflammation and cancer. Once barriers are breached, microbes can elicit proinflammatory or immunosuppressive programmes through various pathways. For example, cancer-associated microbes appear to activate NF- κ B signaling within the tumor microenvironment. Other pattern recognition receptors, such as nucleotide-binding oligomerization domain-like receptor (NLR) family members *NOD-2*, *NLRP3*, *NLRP6* and *NLRP12*, may play a role in mediating colorectal cancer. Likewise *Helicobacter pylori* appears to increase the risk of gastric cancer, due to its driving a chronic inflammatory response in the stomach.

INFLAMMATORY BOWEL DISEASE

Inflammatory bowel disease consists of two different diseases: ulcerative colitis and Crohn's disease and both of these diseases present with disruptions in the gut microbiota (also known as dysbiosis). This dysbiosis presents itself in the form of decreased microbial diversity in the gut, and is correlated to defects in host genes that changes the innate immune response in individuals.

HUMAN IMMUNODEFICIENCY VIRUS

The HIV disease progression influences the composition and function of the gut microbiota, with notable differences between HIV-negative, HIV-positive, and post-ART HIV-positive populations. HIV decreases the integrity of the gut epithelial barrier function by affecting tight junctions. This breakdown allows for translocation across the gut epithelium, which is thought to contribute to increases in inflammation seen in people with HIV.

Vaginal microbiota plays a role in the infectivity of HIV, with an increased risk of infection and transmission when the woman has bacterial vaginosis, a condition characterized by an abnormal balance of vaginal bacteria. The enhanced infectivity is seen with the increase in pro-inflammatory cytokines and CCR5 +CD4+ cells in the vagina. However, a decrease in infectivity is seen with increased levels of vaginal *Lactobacillus*, which promotes an anti-inflammatory condition.

DEATH

With death the microbiome of the living body collapses and a different microbiome named necrobiome establishes itself. Its predictable changes over time are thought to be useful to help determine the time of death.

ENVIRONMENTAL HEALTH

Studies in 2009 questioned whether the decline in biota (including microfauna) as a result of human intervention might impede human health, hospital safety procedures, food product design, and treatments of disease.

Pathogenesis

INTRODUCTION

The pathogenesis of a disease is the biological mechanism (or mechanisms) that leads to the diseased state. The term can also describe the origin and development of the disease, and whether it is acute, chronic, or recurrent.

DESCRIPTION

Types of pathogenesis include microbial infection, inflammation, malignancy and tissue breakdown. For example, bacterial pathogenesis is the mechanism by which bacteria cause infectious illness. Most diseases are caused by multiple processes. For example, certain cancers arise from dysfunction of the immune system (skin tumors and lymphoma after a renal transplant, which requires immunosuppression). The pathogenic mechanisms of a disease (or condition) are set in motion by the underlying causes, which if controlled would allow the disease to be prevented. Often, a potential cause is identified by epidemiological observations before a pathological link can be drawn between the cause and the disease. The pathological perspective can be directly integrated into an epidemiological approach in the interdisciplinary field of molecular pathological epidemiology. Molecular pathological epidemiology can help to assess pathogenesis and causality by means of linking a potential risk factor to molecular pathologic signatures of a disease. Thus, the molecular pathological epidemiology paradigm can advance the area of causal inference.

ACUTE (MEDICINE)

In medicine, describing a disease as acute denotes that it is of short duration and, as a corollary of that, of recent onset. The quantitation of how much time constitutes “short” and “recent” varies by disease and by context, but the core denotation of “acute” is always qualitatively in contrast with “chronic”, which denotes long-lasting disease (for example, in acute leukemia and chronic leukemia). In addition, “acute” also often connotes two other meanings: sudden onset and severity, such as in acute myocardial infarction (EMI), where suddenness and severity are both established aspects of the meaning. It thus often connotes that the condition is fulminant (as in the EMI example), but not always (as in acute rhinitis, which is usually synonymous with the common cold). The one thing that acute MI and acute rhinitis have in common is that they are not chronic. They can happen again (as in recurrent pneumonia, that is, multiple acute pneumonia episodes), but they are not the same case ongoing for months or years (unlike chronic obstructive pulmonary disease, which is). A noncount sense of “acute disease” refers to the acute phase, that is, a short course, of any disease entity. For example, in an article on ulcerative enteritis in poultry, the author says, “in acute disease there may be increased mortality without any obvious signs”, referring to the acute form or phase of ulcerative enteritis.

MEANING VARIATIONS

Not all acute diseases or injuries are severe, and vice versa. For example, a mild stubbed toe is an acute injury. Similarly, many acute upper respiratory infections and acute gastroenteritis cases in adults are mild and usually resolve within a few days or weeks. The term “acute” is also included in the definition of several diseases, such as severe acute respiratory syndrome, acute leukemia, acute myocardial infarction, and acute hepatitis. This is often to distinguish diseases from their chronic forms, such as chronic leukemia, or to highlight the sudden onset of the disease, such as acute myocardial infarct.

Related Terminology

Table. Related terms include.

Term	Meaning
Peracute	Very acute or violent. Denotes fulminant, whereas "acute" only sometimes connotes fulminant. Peracute ("very") is not to be confused with preacute ("before", the opposite of postacute).
Recurrent	"Happening again"—the concept is often one of multiple acute episodes. Relapse can mean the same as recurrent, although relapse is usually used to describe recurrence of chronic conditions that go into remission and then recur.
Acute on chronic	An acute exacerbation of a chronic condition. It is applied to a variety of conditions, including liver failure, subdural hematoma, renal failure, respiratory failure, and bronchitis.
Acute on chronic inflammation	A term sometimes used in pathology to describe a pattern of inflammation which is a mixture of chronic and acute inflammation. It may be seen in asthma, rheumatoid arthritis, chronic peptic ulcer, chronic periodontitis, tuberculosis, tonsillitis and other conditions.
Subacute	A vaguely defined state that is clearly not acute, but rather between acute and chronic, for example subacute endocarditis, or subacute sclerosing panencephalitis.
Chronic	A long-term condition.

ACUTE CARE

Acute care is the early and specialist management of adult patients suffering from a wide range of medical conditions requiring urgent or emergency care usually within 48 hours of admission or referral from other specialties. Acute hospitals are those intended for short-term medical and/or surgical treatment and care. The related medical speciality is acute medicine.

CHRONIC CONDITION

A chronic condition is a human health condition or disease that is persistent or otherwise long-lasting in its effects or a disease that comes with time. The term *chronic* is often applied when the course of the disease lasts for more than three months. Common chronic diseases include arthritis, asthma, cancer, COPD, diabetes and some viral diseases such as hepatitis C and HIV/AIDS. An illness which is lifelong because it ends in death is a terminal illness. In medicine, a chronic condition can be distinguished from one that is acute (recent in onset); additionally, a recurrent condition can relapse repeatedly, with periods of remission in between. The non-communicable diseases are also usually lasting medical conditions but are distinguished by their non-infectious causes. In contrast, some chronic diseases, such as HIV/AIDS, are caused by transmissible infections. In the United States 25 per cent of adults have at least two chronic conditions. Chronic diseases constitute a major cause of mortality, and the World Health Organization (WHO) attributes 38 million deaths a year to non-communicable diseases.

TYPES

Chronic conditions have often been used to describe the various health related states of the human body such as syndromes, physical impairments, disabilities as well as diseases. Epidemiologists have found interest in

chronic conditions due to the fact they contribute to disease, disability, and diminished physical and/or mental capacity. For example, high blood pressure or hypertension is considered to be not only a chronic condition itself but also correlated to diseases such as heart attack or stroke. Additionally, some socioeconomic factors may be considered as a chronic condition as they lead to disability in daily life. An important one that public health officials in the social science setting have begun highlighting is chronic poverty.

The list below includes these chronic conditions and diseases:

In 2015 the World Health Organisation produced a report on non-communicable diseases, citing the four major types as:

- Cancers
- Cardiovascular diseases, including cerebrovascular disease, heart failure, and ischemic cardiopathy
- Chronic respiratory diseases, such as asthma and chronic obstructive pulmonary disease (COPD)
- Diabetes mellitus

Other examples of chronic diseases and health conditions include:

- Addiction
- Alzheimer's disease
- Atrial fibrillation
- Attention deficit hyperactivity disorder
- Autoimmune diseases, such as ulcerative colitis, lupus erythematosus, Crohn's disease, coeliac disease, Hashimoto's thyroiditis, and relapsing polychondritis
- Bipolar disorder
- Blindness
- Cerebral palsy (all types)
- Chronic graft-versus-host disease (GVHD)
- Chronic hepatitis
- Chronic kidney disease
- Chronic osteoarticular diseases, such as osteoarthritis and rheumatoid arthritis
- Chronic pain syndromes, such as post-vasectomy pain syndrome and complex regional pain syndrome
- Depression
- Deafness and hearing impairment
- Eating disorders
- Ehlers–Danlos syndrome (various types)
- Endometriosis
- Epilepsy
- Fibromyalgia
- HIV/AIDS
- Huntington's disease
- Hypertension
- Lyme disease
- Multiple sclerosis
- Myalgic encephalomyelitis (a.k.a. chronic fatigue syndrome)
- Narcolepsy
- Obesity
- Osteoporosis
- Parkinson's disease
- Periodontal disease
- Postural orthostatic tachycardia syndrome

- Sickle cell anemia and other hemoglobin disorders
- Sleep apnea
- Thyroid disease
- Tobacco Use and Related Conditions

RISK FACTORS

While risk factors vary with age and gender, most of the common chronic diseases in the US are caused by dietary, lifestyle and metabolic risk factors that are also responsible for the resulting mortality. Therefore, these conditions might be prevented by behavioral changes, such as quitting smoking, adopting a healthy diet, and increasing physical activity. Social determinants are important risk factors for chronic diseases. Social factors, *e.g.*, socioeconomic status, education level, and race/ethnicity, are a major cause for the disparities observed in the care of chronic disease.

Lack of access and delay in receiving care result in worse outcomes for patients from minorities and underserved populations. Those barriers to medical care complicate patients monitoring and continuity in treatment. In the US, Minorities and low-income populations are less likely to access and receive preventive services necessary to detect conditions at an early stage. The majority of US health care and economic costs associated with medical conditions are for the costs of chronic diseases and conditions and associated health risk behaviors. Eighty-four percent of all health care spending in 2006 was for the 50 per cent of the population who have one or more chronic medical conditions (CDC, 2014).

PREVENTION

A growing body of evidence supports that prevention is effective in reducing the effect of chronic conditions; in particular, early detection results in less severe outcomes. Clinical preventive services include screening for the existence of the disease or predisposition to its development, counseling and immunizations against infectious agents. Despite their effectiveness, the utilization of preventive services is typically lower than for regular medical services. In contrast to their apparent cost in time and money, the benefits of preventive services are not directly perceived by patient because their effects are on the long term or might be greater for society as a whole than at the individual level. Therefore, public health programmes are important in educating the public, and promoting healthy lifestyles and awareness about chronic diseases. While those programmes can benefit from funding at different levels (state, federal, private) their implementation is mostly in charge of local agencies and community-based organizations.

Studies have shown that public health programmes are effective in reducing mortality rates associated to cardiovascular disease, diabetes and cancer, but the results are somewhat heterogeneous depending on the type of condition and the type of programmes involved. For example, results from different approaches in cancer prevention and screening depended highly on the type of cancer. The rising number of patient with chronic diseases has renewed the interest in prevention and its potential role in helping control costs. In 2008, the Trust for America's Health produced a report that estimated investing \$10 per person annually in community-based programmes of proven effectiveness and promoting healthy lifestyle (increase in physical activity, healthier diet and preventing tobacco use) could save more than \$16 billion annually within a period of just five years.

EPIDEMIOLOGY

In the treatment of HIV, the success of antiretroviral therapies means that many patients will experience this infection as a chronic disease that for many will span several decades of their life.

United States

In the United States, nearly one in two Americans (133 million) has at least one chronic medical condition, with most subjects (58 per cent) between the ages of 18 and 64. The number is projected to increase by more than one percent per year by 2030, resulting in an estimated chronically ill population of 171 million. The most common chronic conditions are high blood pressure, arthritis, respiratory diseases like emphysema, and high cholesterol. According to research by the Centers for Disease Control and Prevention, chronic disease is also especially a concern in the elderly population in America. Chronic diseases like stroke, heart disease, and cancer were among the leading causes of death among Americans aged 65 or older in 2002, accounting for 61 per cent of all deaths among this subset of the population. While the majority of chronic conditions are found in individuals between the ages of 18 and 64, it is estimated that at least 80 per cent of older Americans are currently living with some form of a chronic condition, with 50 per cent of this population having two or more chronic conditions.

The two most common chronic conditions in the elderly are high blood pressure and arthritis, with diabetes, coronary heart disease, and cancer also being reported among the elder population. In examining the statistics of chronic disease among the living elderly, it is also important to make note of the statistics pertaining to fatalities as a result of chronic disease. Heart disease is the leading cause of death from chronic disease for adults older than 65, followed by cancer, stroke, diabetes, chronic lower respiratory diseases, influenza and pneumonia, and, finally, Alzheimer's disease. Though the rates of chronic disease differ by race for those living with chronic illness, it is interesting to note that the statistics for leading causes of death among elderly are nearly identical across racial/ethnic groups. Chronic illnesses cause about 70 per cent of deaths in the US and in 2002 chronic conditions (heart disease, cancers, stroke, chronic respiratory diseases, diabetes, Alzheimer's disease, mental illness and kidney diseases) were 6 of the top ten causes of mortality in the general US population. In the United States, 90 per cent of seniors have at least one chronic disease, and 77 per cent have two or more chronic conditions.

ECONOMIC IMPACT

United States

Chronic diseases are a major factor in the continuous growth of medical care spending. Healthy People 2010 reported that more than 75 per cent of the \$2 trillion spent annually in US medical care are due to chronic conditions; spending are even higher in proportion for Medicare beneficiaries (aged 65 years and older). Spending growth is driven in part by the greater prevalence of chronic illnesses, and the longer life expectancy of the population. Also improvement in treatments has significantly extended the life spans of patients with chronic diseases but results in additional costs over long period of time. A striking success is the development of combined antiviral therapies that led to remarkable improvement in survival rates and quality of life of HIV-infected patients.

In addition to direct costs in health care, chronic diseases are a significant burden to the economy, through limitations in daily activities, loss in productivity and loss of days of work. A particular concern is the rising rates of overweight and obesity in all segments of the US population. Obesity itself is a medical condition and not a disease, but it constitutes a major risk factor for developing chronic illnesses, such as diabetes, stroke, cardiovascular disease and cancers. Obesity results in significant health care spending and indirect costs, as illustrated by a recent study from the Texas comptroller reporting that obesity alone cost Texas businesses an extra \$9.5 billion in 2009, including more than \$4 billion for health care, \$5 billion for lost productivity and absenteeism, and \$321 million for disability.

RELAPSE

In medicine, relapse or recidivism is a recurrence of a past (typically medical) condition. For example, multiple sclerosis and malaria often exhibit peaks of activity and sometimes long periods of dormancy, followed by relapse or recrudescence. In the context of drug use, relapse or reinstatement of drug-seeking behaviour, is a form of spontaneous recovery that involves the recurrence of pathological drug use after a period of abstinence. Relapse is often observed in individuals who have developed a drug addiction or either form of drug dependence.

CAUSES

Substances that may cause addiction or dependence include:

- Stimulants increase activity in the cerebral cortex leading to increased motor activity.
 - (a) Cocaine
 - (b) Amphetamine
 - (c) Methamphetamine
- Depressants slow down neuronal activity.
 - (a) Benzodiazepines (*i.e.*: Xanax)
 - (b) Opioids activate or block opioid receptors in the brain typically to reduce the effect of pain. Some common opioids are:
 - Morphine
 - Methaone
 - Opium
 - Heroin
 - (c) Alcohol produces disinhibition in the nervous system when introduced, and it depresses the frontal cortex while speeding up the rest of the brain. This can lead to decreased ability to perceive and evaluate risks and make good decisions, and other characteristics of what is commonly known as intoxication.
- Nicotine is neither a stimulant nor a depressant but rather a chemical that is absorbed by the skin and mucous membranes and activates the nicotinic acetylcholine receptors.

DEA Schedules of Controlled Substances

The Drug Enforcement Administration (DEA) has categorized controlled substances into 5 major categories based on the drug's intended use and potential for addiction and dependence. Drugs with the highest addictive potential are listed in DEA Schedules I and II. Schedule I drugs are those with no accepted medical or therapeutic use whereas Schedule II drugs are those that can be used therapeutically but may lead to severe physical or psychological dependence.

Addictive Potential

The addictive potential of a drug varies greatly between substances and is based on both the rewarding and reinforcing properties of a drug. Addictive drugs tend to induce euphoria, otherwise known as a drug high.

Categories of Harm

The three main categories of harm are the physical harm of the drug to the user, the drug's tendency to cause dependence, and effects of the drug on society.

Physical Harm

The first category can be further divided into three parameters of harm: acute physical harm, chronic physical harm, and intravenous harm. Acute harm is defined as the immediate effects associated with use of the given drug such as respiratory depression or myocardial infarction. Chronic harm is the consequence of continued and repeated use such as psychosis or lung disease. Lastly, intravenous harm refers to problems associated with the route of administration such as the spread of blood-borne pathogens like HIV.

Dependence

The second parameter of harm is subdivided into two categories: the physical dependence and psychological dependence.

Social Harm

This category is subdivided into the last three parameters: intoxication, other social harms, and health care costs. These parameters attempt to rate a drug's impact on families, communities, and societies. The nine parameters used to evaluate addictive potential of various substances:

Category of harm	Parameter	Subgroup
Physical harm	1	Acute
	2	Chronic
	3	Intravenous harm
Dependence	4	Intensity of pleasure
	5	Psychological dependence
	6	Physical dependence
Social harm	7	Intoxication
	8	Other social harms
	9	Health-care costs

RISK FACTORS

Dopamine D2 Receptor Availability

The availability of the dopamine receptor D2 plays a role in self-administration and the reinforcing effects of cocaine and other stimulants. The D2 receptor availability has an inverse relationship to vulnerability to the reinforcing effects of the drug. That is, as D2 receptors become limited the user becomes more susceptible to the reinforcing effects of cocaine. It is currently unknown if a predisposition to low D2 receptor availability is possible; however, most studies support the idea that changes in D2 receptor availability are a result, rather than a precursor, of cocaine use. It has also been noted that D2 receptors may return to the level existing prior to drug exposure during long periods of abstinence, a fact which may have implications in relapse treatment.

Social Hierarchy

Social interactions, such as the formation of linear dominance hierarchies, also play a role in vulnerability to drug abuse. Animal studies suggest that there exists a difference in D2 receptor availability between dominant and subordinate animals within a social hierarchy as well as a difference in the function of cocaine to reinforce self-administration in these animal groups. Socially dominant animals exhibit higher availability of D2 receptors and fail to maintain self-administration.

TRIGGERS

Drug taking and relapse are heavily influenced by a number of factors including the pharmacokinetics, dose, and neurochemistry of the drug itself as well as the drug taker's environment and drug-related history. Reinstatement of drug use after a period of non-use or abstinence is typically initiated by one or a combination of the three main triggers: stress, re-exposure to the drug or drug-priming, and environmental cues. These factors may induce a neurochemical response in the drug taker that mimics the drug and thus triggers reinstatement. These cues may lead to a strong desire or intention to use the drug, a feeling termed craving by Abraham Wikler in 1948. The propensity for craving is heavily influenced by all three triggers to relapse and is now an accepted hallmark of substance dependence. Stress is one of the most powerful stimuli for reinstating drug use because stress cues stimulate craving and drug-seeking behaviour during abstinence. Stress-induced craving is also predictive of time to relapse. Comparably, addicted individuals show an increased susceptibility to stressors than do non-addicted controls. Examples of stressors that may induce reinstatement include emotions of fear, sadness, or anger, a physical stressor such as a footshock or elevated sound level, or a social event. Drug-priming is exposing the abstinent user to the drug of abuse, which will induce reinstatement of the drug-seeking behaviour and drug self-administration. Stimuli that have a pre-existing association with a given drug or with use of that drug can trigger both craving and reinstatement. These cues include any items, places, or people associated with the drug.

TREATMENT

Relapse treatment is somewhat of a misnomer because relapse itself is a treatment failure; however there exist three main approaches that are currently used to reduce the likelihood of drug relapse. These include pharmacotherapy, cognitive behavioral techniques, and contingency management. The main goals of treating substance dependence and preventing relapse are to identify the needs that were previously met by use of the drug and to develop the skills needed to meet those needs in an alternative way.

Pharmacotherapy

Drug rehabilitation (often drug rehab or just rehab) is the processes of medical or psychotherapeutic treatment for dependency on psychoactive substances such as alcohol, prescription drugs, and street drugs such as cocaine, heroin or amphetamines. The general intent is to enable the patient to confront substance dependence, if present, and cease substance abuse to avoid the psychological, legal, financial, social, and physical consequences that can be caused, especially by extreme abuse. Treatment includes medication for depression or other disorders, counseling by experts and sharing of experience with other addicts.

Psychological Dependency

Psychological dependency is addressed in many drug rehabilitation programmes by attempting to teach the patient new methods of interacting in a drug-free environment. In particular, patients are generally encouraged, or possibly even required, to not associate with friends who still use the addictive substance. Twelve-step programmes encourage addicts not only to stop using alcohol or other drugs, but to examine and change habits related to their addictions. Many programmes emphasize that recovery is a permanent process without culmination. For legal drugs such as alcohol, complete abstention—rather than attempts at moderation, which may lead to relapse—is also emphasized (“One is too many, and a thousand is never enough.”) Whether moderation is achievable by those with a history of abuse remains a controversial point, but is generally considered unsustainable.

Types of Treatment

The brain's chemical structure is impacted by drugs of abuse and these changes are present long after an individual stops using. This change in brain structure increases risk for relapse, making treatment an important part of the rehabilitation process. Various types of programmes offer help in drug rehabilitation, including: residential treatment (in-patient/ out-patient), local support groups, extended care centers, recovery or sober houses, addiction counselling, mental health, and medical care. Some rehab centers offer age- and gender-specific programmes. In a survey of treatment providers from three separate institutions (the National Association of Alcoholism and Drug Abuse Counselors, Rational Recovery Systems and the Society of Psychologists in Addictive Behaviors) measuring the treatment provider's responses on the Spiritual Belief Scale (a scale measuring belief in the four spiritual characteristics AA identified by Ernest Kurtz); the scores were found to explain 41 per cent of the variance in the treatment provider's responses on the Addiction Belief Scale (a scale measuring adherence to the disease model or the free-will model addiction). Scientific research since 1970 shows that effective treatment addresses the multiple needs of the patient rather than treating addiction alone. In addition, medically assisted drug detoxification or alcohol detoxification alone is ineffective as a treatment for addiction. The National Institute on Drug Abuse (NIDA) recommends detoxification followed by both medication (where applicable) and behavioral therapy, followed by relapse prevention. According to NIDA, effective treatment must address medical and mental health services as well as follow-up options, such as community or family based recovery support systems. Whatever the methodology, patient motivation is an important factor in treatment success.

For individuals addicted to prescription drugs, treatments tend to be similar to those who are addicted to drugs affecting the same brain systems. Medication like methadone and buprenorphine can be used to treat addiction to prescription opiates, and behavioral therapies can be used to treat addiction to prescription stimulants, benzodiazepines, and other drugs.

Types of behavioral therapy include:

- Cognitive-behavioral therapy, which seeks to help patients to recognize, avoid and cope with situations in which they are most likely to relapse.
- Multidimensional family therapy, which is designed to support recovery of the patient by improving family functioning.
- Motivational interviewing, which is designed to increase patient motivation to change behaviour and enter treatment.
- Motivational incentives, which uses positive reinforcement to encourage abstinence from the addictive substance.

Treatment can be a long process and the duration is dependent upon the patient's needs and history of abuse. Research has shown that most patients need at least 3 months of treatment and longer durations are associated with better outcomes.

Medications

Certain opioid medications such as methadone and more recently buprenorphine (In America, "Subutex" and "Suboxone") are widely used to treat addiction and dependence on other opioids such as heroin, morphine or oxycodone. Methadone and buprenorphine are maintenance therapies intended to reduce cravings for opiates, thereby reducing illegal drug use, and the risks associated with it, such as disease, arrest, incarceration, and death, in line with the philosophy of harm reduction. Both drugs may be used as maintenance medications (taken for an indefinite period of time), or used as detoxification aids. All available studies collected in the 2005 Australian National Evaluation of Pharmacotherapies for Opioid Dependence suggest that maintenance treatment is preferable, with very high rates (79–100 per cent) of relapse within three months of detoxification from

LAAM, buprenorphine, and methadone. According to the National Institute on Drug Abuse (NIDA), patients stabilized on adequate, sustained doses of methadone or buprenorphine can keep their jobs, avoid crime and violence, and reduce their exposure to HIV and Hepatitis C by stopping or reducing injection drug use and drug-related high risk sexual behaviour. Naltrexone is a long-acting opioid antagonist with few side effects. It is usually prescribed in outpatient medical conditions. Naltrexone blocks the euphoric effects of alcohol and opiates. Naltrexone cuts relapse risk during the first 3 months by about 36 per cent. However, it is far less effective in helping patients maintain abstinence or retaining them in the drug-treatment system (retention rates average 12 per cent at 90 days for naltrexone, average 57 per cent at 90 days for buprenorphine, average 61 per cent at 90 days for methadone).

Ibogaine is a hallucinogenic drug promoted by certain fringe groups to interrupt both physical dependence and psychological craving to a broad range of drugs including narcotics, stimulants, alcohol and nicotine. To date, there have never been any controlled studies showing it to be effective, and it is accepted as a treatment by no association of physicians, pharmacists, or addictionologists. There have been several deaths related to ibogaine use, which causes tachycardia and long QT syndrome. The drug is an illegal Schedule I controlled substance in the United States, and the foreign facilities in which it is administered tend to have little oversight, and range from motel rooms to one moderately-sized rehabilitation center. A few antidepressants have been proven to be helpful in the context of smoking cessation/nicotine addiction, these medications include bupropion and nortriptyline. Bupropion inhibits the reuptake of norepinephrine and dopamine and has been FDA approved for smoking cessation, while nortriptyline is a tricyclic antidepressant which has been used to aid in smoking cessation though it has not been FDA approved for this indication. Acamprosate, disulfiram and topiramate (a novel anticonvulsant sulphonated sugar) are also used to treat alcohol addiction. Acamprosate has shown effectiveness for patients with severe dependence, helping them to maintain abstinence for several weeks or months. Disulfiram (also called Antabuse) produces a very unpleasant reaction when drinking alcohol that includes flushing, nausea and palpitations. It is more effective for patients with high motivation and some addicts use it only for high risk situations. Patients who wish to continue drinking or may be likely to relapse, should not take disulfiram as it can result in the disulfiram-alcohol reaction mentioned previously, which is very serious and can even be fatal. Nitrous oxide, also sometimes known as laughing gas, is a legally available gas used for purposes that include anesthesia during certain dental and surgical procedures, as well as food preparation and the fueling of rocket and racing engines. Substance abusers also sometimes use the gas as an inhalant.

Like all other inhalants, it's popular because it provides consciousness-altering effects while allowing users to avoid some of the legal issues surrounding illicit or illegal drugs of abuse. Abuse of nitrous oxide can produce significant short-term and long-term damage to human health, including a form of oxygen starvation called hypoxia, brain damage, and a serious vitamin B12 deficiency that can lead to nerve damage. Although dangerous and addictive in its own right, nitrous oxide has been shown to be an effective treatment for a number of addictions.

Residential Treatment

In-patient residential treatment for alcohol abuse is usually quite expensive without proper insurance. Most American programmes follow a traditional 28–30 day programme length. The length is based solely upon providers' experience in the 1940s that clients needed about one week to get over the physical changes, another week to understand the programme, and another week or two to become stable. 70 to 80 percent of American residential alcohol treatment programmes provide 12-step support services. These include, but are not limited to AA, NA, CA, Al-Anon. One recent study suggests the importance of family participation in residential treatment patient retention, finding "increased programme completion rate for those with a family member or significant other involved in a seven-day family programme."

Experimental Treatment

The Nature of Things, a CBC Television programme by David Suzuki, explored an experimental drug treatment by Dr. Gabor Maté in which the substance Ayahuasca was used to treat addicts in Vancouver.

Recovery

The definition of recovery remains divided and subjective in drug rehabilitation, as there are no set standards for measuring recovery. The Betty Ford Institute defined recovery as achieving complete abstinence as well as personal wellbeing while other studies have considered “near abstinence” as a definition. The wide range of meanings has complicated the process of choosing rehabilitation programmes.

Criminal Justice

Drug rehabilitation is sometimes part of the criminal justice system. People convicted of minor drug offenses may be sentenced to rehabilitation instead of prison, and those convicted of driving while intoxicated are sometimes required to attend Alcoholics Anonymous meetings. There are a number of ways to address an alternative sentence in a drug possession or DUI case; increasingly, American courts are willing to explore outside-the-box methods for delivering this service. There have been lawsuits filed, and won, regarding the requirement of attending Alcoholics Anonymous and other twelve-step meetings as being inconsistent with the Establishment Clause of the First Amendment of the U. S. Constitution, mandating separation of church and state. In some cases, individuals can be court ordered to drug rehabilitation by the state through legislation like the Marchman Act.

Counseling

Traditional addiction treatment is based primarily on counseling.

Counselors help individuals identifying behaviors and problems related to their addiction. It can be done on an individual basis, but it's more common to find it in a group setting and can include crisis counseling, weekly or daily counseling, and drop-in counseling supports. They are trained to develop recovery programmes that help to reestablish healthy behaviors and provide coping strategies whenever a situation of risk happens. It's very common to see them work also with family members who are affected by the addictions of the individual, or in a community to prevent addiction and educate the public. Counselors should be able to recognize how addiction affects the whole person and those around him or her. Counseling is also related to “Intervention”; a process in which the addict's family requests help from a professional to get this person into drug treatment. This process begins with one of this professionals' first goals: breaking down denial of the person with the addiction. Denial implies lack of willingness from the patients or fear to confront the true nature of the addiction and to take any action to improve their lives, besides of continuing the destructive behaviour. Once this has been achieved, professional coordinates with the addict's family to support them on getting this family member to alcohol drug rehabilitation immediately, with concern and care for this person. Otherwise, this person will be asked to leave and expect no support of any kind until going into drug rehabilitation or alcoholism treatment. An intervention can also be conducted in the workplace environment with colleagues instead of family. One approach with limited applicability is the Sober Coach. In this approach, the client is serviced by provider(s) in his or her home and workplace – for any efficacy, around-the-clock – who functions much like a nanny to guide or control the patient's behaviour.

Twelve-step Programmes

The disease model of addiction has long contended the maladaptive patterns of alcohol and substance use displayed by addicted individuals are the result of a lifelong disease that is biological in origin and exacerbated

by environmental contingencies. This conceptualization renders the individual essentially powerless over his or her problematic behaviors and unable to remain sober by himself or herself, much as individuals with a terminal illness are unable to fight the disease by themselves without medication. Behavioral treatment, therefore, necessarily requires individuals to admit their addiction, renounce their former lifestyle, and seek a supportive social network who can help them remain sober. Such approaches are the quintessential features of Twelve-step programmes, originally published in the book *Alcoholics Anonymous* in 1939. These approaches have met considerable amounts of criticism, coming from opponents who disapprove of the spiritual-religious orientation on both psychological and legal grounds. Nonetheless, despite this criticism, outcome studies have revealed that affiliation with twelve-step programmes predicts abstinence success at 1-year follow-up for alcoholism. Different results have been reached for other drugs, with the twelve steps being less beneficial for addicts to illicit substances, and least beneficial to those addicted to the physiologically and psychologically addicting opioids, for which maintenance therapies are the gold standard of care.

Client-centered Approaches

In his influential book, *Client-Centered Therapy*, in which he presented the client-centered approach to therapeutic change, psychologist Carl Rogers proposed there are three necessary and sufficient conditions for personal change: unconditional positive regard, accurate empathy, and genuineness. Rogers believed the presence of these three items in the therapeutic relationship could help an individual overcome any troublesome issue, including alcohol abuse. To this end, a 1957 study compared the relative effectiveness of three different psychotherapies in treating alcoholics who had been committed to a state hospital for sixty days: a therapy based on two-factor learning theory, client-centered therapy, and psychoanalytic therapy. Though the authors expected the two-factor theory to be the most effective, it actually proved to be deleterious in outcome. Surprisingly, client-centered therapy proved most effective. It has been argued, however, these findings may be attributable to the profound difference in therapist outlook between the two-factor and client-centered approaches, rather than to client-centered techniques per se. The authors note two-factor theory involves stark disapproval of the clients' "irrational behaviour" (p. 350); this notably negative outlook could explain the results. A variation of Rogers' approach has been developed in which clients are directly responsible for determining the goals and objectives of the treatment. Known as Client-Directed Outcome-Informed therapy (CDOI), this approach has been utilized by several drug treatment programmes, such as Arizona's Department of Health Services.

Psychoanalysis

Psychoanalysis, a psychotherapeutic approach to behaviour change developed by Sigmund Freud and modified by his followers, has also offered an explanation of substance abuse. This orientation suggests the main cause of the addiction syndrome is the unconscious need to entertain and to enact various kinds of homosexual and perverse fantasies, and at the same time to avoid taking responsibility for this. It is hypothesised specific drugs facilitate specific fantasies and using drugs is considered to be a displacement from, and a concomitant of, the compulsion to masturbate while entertaining homosexual and perverse fantasies. The addiction syndrome is also hypothesised to be associated with life trajectories that have occurred within the context of traumatogenic processes, the phases of which include social, cultural and political factors, encapsulation, traumatophilia, and masturbation as a form of self-soothing. Such an approach lies in stark contrast to the approaches of social cognitive theory to addiction—and indeed, to behaviour in general—which holds human beings regulate and control their own environmental and cognitive environments, and are not merely driven by internal, driving impulses. Additionally, homosexual content is not implicated as a necessary feature in addiction.

Relapse prevention:

- An influential cognitive-behavioral approach to addiction recovery and therapy has been Alan Marlatt's (1985) Relapse Prevention approach. Marlatt describes four psychosocial processes relevant to the

addiction and relapse processes: self-efficacy, outcome expectancies, attributions of causality, and decision-making processes. Self-efficacy refers to one's ability to deal competently and effectively with high-risk, relapse-provoking situations. Outcome expectancies refer to an individual's expectations about the psychoactive effects of an addictive substance. Attributions of causality refer to an individual's pattern of beliefs that relapse to drug use is a result of internal, or rather external, transient causes (*e.g.*, allowing oneself to make exceptions when faced with what are judged to be unusual circumstances). Finally, decision-making processes are implicated in the relapse process as well. Substance use is the result of multiple decisions whose collective effects result in consumption of the intoxicant. Furthermore, Marlatt stresses some decisions—referred to as apparently irrelevant decisions—may seem inconsequential to relapse, but may actually have downstream implications that place the user in a high-risk situation.

- For example: As a result of heavy traffic, a recovering alcoholic may decide one afternoon to exit the highway and travel on side roads. This will result in the creation of a high-risk situation when he realizes he is inadvertently driving by his old favorite bar. If this individual is able to employ successful coping strategies, such as distracting himself from his cravings by turning on his favorite music, then he will avoid the relapse risk (PATH 1) and heighten his efficacy for future abstinence. If, however, he lacks coping mechanisms—for instance, he may begin ruminating on his cravings (PATH 2)—then his efficacy for abstinence will decrease, his expectations of positive outcomes will increase, and he may experience a lapse—an isolated return to substance intoxication. So doing results in what Marlatt refers to as the Abstinence Violation Effect, characterized by guilt for having gotten intoxicated and low efficacy for future abstinence in similar tempting situations. This is a dangerous pathway, Marlatt proposes, to full-blown relapse.

Cognitive therapy:

- An additional cognitively-based model of substance abuse recovery has been offered by Aaron Beck, the father of cognitive therapy and championed in his 1993 book *Cognitive Therapy of Substance Abuse*. This therapy rests upon the assumption addicted individuals possess core beliefs, often not accessible to immediate consciousness (unless the patient is also depressed). These core beliefs, such as “I am undesirable,” activate a system of addictive beliefs that result in imagined anticipatory benefits of substance use and, consequentially, craving. Once craving has been activated, permissive beliefs (“I can handle getting high just this one more time”) are facilitated. Once a permissive set of beliefs have been activated, then the individual will activate drug-seeking and drug-ingesting behaviors. The cognitive therapist's job is to uncover this underlying system of beliefs, analyze it with the patient, and thereby demonstrate its dysfunctionality. As with any cognitive-behavioral therapy, homework assignments and behavioral exercises serve to solidify what is learned and discussed during treatment.

Emotion Regulation and Mindfulness

A growing literature is demonstrating the importance of emotion regulation in the treatment of substance abuse. Considering that nicotine and other psychoactive substances such as cocaine activate similar psychopharmacological pathways, an emotion regulation approach may be applicable to a wide array of substance abuse. Proposed models of affect-driven tobacco use have focused on negative reinforcement as the primary driving force for addiction; according to such theories, tobacco is used because it helps one escape from the undesirable effects of nicotine withdrawal or other negative moods. Acceptance and commitment therapy (ACT), is showing evidence that it is effective in treating substance abuse, including the treatment of poly-substance abuse and cigarette smoking. Mindfulness programmes that encourage patients to be aware of their own experiences in the present moment and of emotions that arise from thoughts, appear

to prevent impulsive/compulsive responses. Research also indicates that mindfulness programmes can reduce the consumption of substances such as alcohol, cocaine, amphetamines, marijuana, cigarettes and opiates.

Behavioral Models

Behavioral models make use of principles of functional analysis of drinking behaviour. Behaviour models exists for both working with the substance abuser (Community Reinforcement Approach) and their family (Community Reinforcement Approach and Family Training). Both these models have had considerable research success for both efficacy and effectiveness. This model lays much emphasis on the use of problem solving techniques as a means of helping the addict to overcome his/her addiction.

Criticism

Despite ongoing efforts to combat addiction, there has been evidence of clinics billing patients for treatments that may not guarantee their recovery. This is a major problem as there are numerous claims of fraud in drug rehabilitation centers, where these centers are billing insurance companies for under delivering much needed medical treatment while exhausting patients' insurance benefits. In California, there are movements and law regarding this matter, particularly the California Insurance Fraud Prevention Act (IFPA) which declares it unlawful to unknowingly conduct such businesses. Under the Affordable Care Act and the Mental Health Parity Act, rehabilitation centers are able to bill insurance companies for substance abuse treatment. With long waitlists in limited state funded rehabilitation centers, controversial private centers rapidly emerged. One popular model, known as the Florida Model for rehabilitation centers, is often criticized for fraudulent billing to insurance companies. Under the guise of helping patients with opioid addiction, these centers would offer addicts free rent or up to \$500 per month to stay in their "sober homes", then charge insurance companies as high as \$5,000 to \$10,000 per test for simple urine tests. Little attention is paid to patients in terms of addiction intervention as these patients have often been known to continue drug use during their stay in these centers. Since 2015, these centers have been under federal and state criminal investigation. As of 2017 in California, there are only 16 investigators in the CA Department of Health Care Services investigating over 2,000 licensed rehab centers.

Cognitive Behavioral Techniques

The various behavioral approaches to treating relapse focus on the precursors and consequences of drug taking and reinstatement. Cognitive behavioral techniques (CBT) incorporate Pavlovian conditioning and operant conditioning, characterized by positive reinforcement and negative reinforcement, in order to alter the cognitions, thoughts, and emotions associated with drug taking behaviour. A main approach of CBT is cue exposure, during which the abstinent user is repeatedly exposed to the most salient triggers without exposure to the substance in hopes that the substance will gradually lose the ability to induce drug-seeking behaviour. This approach is likely to reduce the severity of a relapse than to prevent one from occurring altogether. Another method teaches addicts basic coping mechanisms to avoid using the illicit drug. It is important to address any deficits in coping skills, to identify the needs that likely induce drug-seeking, and to develop another way to meet them.

Relapse Prevention

Relapse prevention (RP) is a cognitive-behavioral approach to relapse with the goal of identifying and preventing high-risk situations such as substance abuse, obsessive-compulsive behaviour, sexual offending, obesity, and depression. It is an important component in the treatment process for alcoholism, or alcohol dependence.

Underlying Assumptions

Relapse is seen as both an outcome and a transgression in the process of behaviour change. An initial setback or lapse may translate into either a return to the previous problematic behaviour, known as relapse, or the individual turning again towards positive change, called prolapse. A relapse often occurs in the following stages: emotional relapse, mental relapse, and finally, physical relapse. Each stage is characterized by feelings, thoughts, and actions that ultimately lead to the individual returning to their old behaviour. Relapse is thought to be multi-determined, especially by self-efficacy, outcome expectancies, craving, motivation, coping, emotional states, and interpersonal factors. In particular, high self-efficacy, negative outcome expectancies, potent availability of coping skills following treatment, positive affect, and functional social support are expected to predict positive outcome. Craving has not historically been shown to serve as a strong predictor of relapse.

Efficacy and Effectiveness

Carroll *et al.* conducted a review of 24 other trials and concluded that RP was more effective than no treatment and was equally effective as other active treatments such as supportive psychotherapy and interpersonal therapy in improving substance use outcomes. Irvin and colleagues also conducted a meta-analysis of RP techniques in the treatment of alcohol, tobacco, cocaine, and polysubstance use, and upon reviewing 26 studies, concluded that RP was successful in reducing substance use and improving psychosocial adjustment. RP seemed to be most effective for individuals with alcohol problems, suggesting that certain characteristics of alcohol use are amenable to the RP. Miller *et al.* (1996) found the GORSKI/CENAPS relapse warning signs to be a good predictor of the occurrence of relapse on the AWARE scale ($r = .42$, $p < .001$).

Systems Theory

Some theorists, including Katie Witkiewitz and G. Alan Marlatt, borrowing ideas from systems theory, conceptualize relapse as a multidimensional, complex system. Such a nonlinear dynamical system is believed to be able to best predict the data witnessed, which commonly includes cases where small changes introduced into the equation seem to have large effects. The model also introduces concepts of self-organization, feedback loops, timing/context effects, and interplay between tonic and phasic processes. Rami Jumnoodoo and Dr. Patrick Coyne, in London UK, have been working with National Health Service users and carers over the past ten years to transfer RP theory into the field of adult mental health. The uniqueness of the model is the sustainment of change by developing service users and carers as 'experts' - following RP as an educational process and graduating as Relapse Prevention Practitioners. The work has won many national awards, been presented at many conferences, and has resulted in a number of publications. Terence Gorski MA has developed the CENAPS (Center for Applied Science) model for relapse prevention including Relapse Prevention Counseling (Gorski, Counseling For Relapse Prevention, 1983) and a system for certification of Relapse Prevention Specialists (CRPS).

Contingency Management

Contingency management (CM) is most-widely used in the field of substance abuse, often implemented as part of clinical behaviour analysis. CM refers to the application of the three-term contingency (or operant conditioning), which uses stimulus control and positive reinforcement to change behaviour. Patients' behaviors are rewarded (or, less often, punished); generally, adherence to or failure to adhere to programme rules and regulations or their treatment plan. CM derives from the science of applied behaviour analysis (ABA), and by most evaluations, its procedures produces one of the largest effect sizes out of all mental health and educational interventions.

Token Economies

One form of contingency management is the token economy system. Token systems can be used in an individual or group format. Token systems have been shown to be successful with a diverse array of populations including those suffering from addiction, those with special needs, and delinquents. However, recent research questions the use of token systems with very young children. The exception to the last would be the treatment of stuttering. The goal of such systems is to gradually thin out and to help the person begin to access the natural community of reinforcement (the reinforcement typically received in the world for performing the behaviour). Walker (1990) presents an overview of token systems and combining such procedures with other interventions in the classroom. He relates the comprehensiveness of token systems to the child's level of difficulty.

Voucher Programmes and Related Applications in Addiction Treatment

Another form of contingency management is voucher programmes. In voucher-based contingency management patients earn vouchers exchangeable for retail items contingent upon objectively verified abstinence from recent drug use or compliance with other behaviour-change targets. This particular form of contingency management was introduced in the early 1990s as a treatment for cocaine dependence. The approach is the most reliably effective method for producing cocaine abstinence in controlled clinical trials. Medication take-home privileges is another form of contingency management frequently used in methadone maintenance treatment. Patients are permitted to “earn” take-home doses of their methadone in exchange for increasing, decreasing, or ceasing certain behaviors. For example, a patient may be given one take-home dose per week after submitting negative drug screens (generally via urine testing) for three months. (It is worth noting that take home-doses (or “bottles”) are seen as desirable rewards because they allow patients to come to the clinic less often to obtain their medication). Based in applied behaviour analysis (ABA), contingency management includes techniques such as choice, shaping, time-out, making contracts between the therapist and patient, and token economy. Contingent vouchers are also used to cease smoking addictions. One study claims that addicts with substance use disorders can receive help with their addiction through the use of voucher-based treatment for smoking. In addition, nicotine replacement (NRT) can help with addiction combined with the vouchers.

Level Systems

Level systems are often employed as a form of contingency management system. Level systems are designed such that once one level is achieved, then the person earns all the privileges for that level and the levels lower than it.

Effectiveness in Addiction Programmes

A recent meta-analysis of contingency management in drug programmes showed that it has a large effect. These contingencies are delivered based on abstinence and attendance goals and can take the form of vouchers, the opportunity to win prizes, or privileges. They have been used with single problem addictions as well as dual diagnoses and homeless. Overall contingency management has been found to be an effective and cost efficient addition to drug treatment. In contrast to these findings in a recent study, the researchers found out that Nicotine replacement treatment only improved the effects of contingent vouchers on short-term smoking abstinence. However, in the long term the effects of contingent vouchers had no impact on tobacco resistance.

Organizations

Many organizations exist for board certified behaviour analysts using contingency management around the world.

ANIMAL MODELS

There are vast ethical limitations in drug addiction research because humans cannot be allowed to self-administer drugs for the purpose of being studied. However, much can be learned about drugs and the neurobiology of drug taking by the examination of laboratory animals. Most studies are performed on rodents or non-human primates with the latter being most comparable to humans in pharmacokinetics, anatomy of the prefrontal cortex, social behaviour, and life span. Other advantages to studying relapse in non-human primates include the ability of the animal to reinstate self-administration, and to learn complex behaviors in order to obtain the drug. Animal studies have shown that a reduction in negative withdrawal symptoms is not necessary to maintain drug taking in laboratory animals; the key to these studies is operant conditioning and reinforcement.

Protocols

Self-administration

To self-administer the drug of interest the animal is implanted with an intravenous catheter and seated in a primate chair equipped with a response lever. The animal is seated in a ventilated chamber and trained on a schedule of drug self-administration.

In many studies the self-administration task begins with presentation of a stimulus light (located near the response panel) that may change colors or turn off upon completion of the operant task. The change in visual stimulus is accompanied by an injection of the given drug through the implanted catheter. This schedule is maintained until the animals learn the task.

Extinction

Extinction in non-human primates is analogous, with some limitations, to abstinence in humans. In order to extinguish drug-seeking behaviour the drug is substituted with a saline solution. When the animal performs the task it has been trained to perform it is no longer reinforced with an injection of the drug. The visual stimulus associated with the drug and completion of the task is also removed. The extinction sessions are continued until the animal ceases the drug-seeking behaviour by pressing the lever.

Reinstatement

After the animal's drug-seeking behaviour is extinguished, a stimulus is presented to promote the reinstatement of that same drug-seeking behaviour (*i.e.*, relapse). For example, if the animal receives an injection of the drug in question it will likely begin working on the operant task for which it was previously reinforced. The stimulus may be the drug itself, the visual stimulus that was initially paired with the drug intake, or a stressor such as an acoustic startle or foot shock.

Neuroimaging

Neuroimaging has contributed to the identification of the neural components involved in drug reinstatement as well as drug-taking determinants such as the pharmacokinetics, neurochemistry, and dose of the drug. The neuroimaging techniques used in non-human primates include positron emission tomography (PET), which uses radiolabeled ligand tracers to measure neurochemistry in vivo and single-photon emission computed tomography (SPECT). Functional magnetic resonance imaging (fMRI) is widely used in human subjects because it has much higher resolution and eliminates exposure to radiation.

Limitations

Although the reinstatement protocols are used frequently in laboratory settings there are some limitations to the validity of the procedures as a model of craving and relapse in humans. The primary limiting factor is that in humans, relapse rarely follows the strict extinction of drug-seeking behaviour. Additionally, human self-reports show that drug-associated stimuli play a lesser role in craving in humans than in the laboratory models. The validity of the model can be examined in three ways: formal equivalence, correlational models, and functional equivalence. There is moderate formal equivalence, or face validity, meaning that the model somewhat resembles relapse as it occurs outside of the laboratory setting; however, there is little face validity for the procedures as a model of craving. The predictive validity, which is assessed by correlational models, has yet to be determined for the procedures. There is sound functional equivalence for the model, which suggests that relapse in the laboratory is reasonably similar to that in nature. Further research into other manipulations or reinforcements that could limit drug taking in non-human primates would be extremely beneficial to the field.

DIFFERENCES BETWEEN SEXES

There exists a higher rate of relapse, shorter periods of abstinence, and higher responsiveness to drug-related cues in women as compared to men. One study suggests that the ovarian hormones, estradiol and progesterone, that exist in females at fluctuating levels throughout the menstrual cycle (or estrous cycle in rodents), play a significant role in drug-primed relapse. There is a marked increase in progesterone levels and a decrease in estradiol levels during the luteal phase. Anxiety, irritability, and depression, three symptoms of both withdrawal and the human menstrual cycle, are most severe in the luteal phase. Symptoms of withdrawal not associated with the cycle, such as hunger, are also enhanced during the luteal phase, which suggests the role of estradiol and progesterone in enhancing symptoms above the naturally occurring level of the menstrual cycle. The symptoms of craving also increase during the luteal phase in humans (it is important to note that the opposite result occurs in female subjects with cocaine addiction suggesting that cyclic changes may be specific for different drugs of abuse). Further, the drug-primed response is reduced during the luteal phase suggesting a time in the cycle during which the urge to continue use may be reduced. These findings implicate a cyclic, hormone-based timing for quitting a drug of abuse and preparing for magnified symptoms of withdrawal or susceptibility to relapse.

Inflammation is part of the complex biological response of body tissues to harmful stimuli, such as pathogens, damaged cells, or irritants, and is a protective response involving immune cells, blood vessels, and molecular mediators.

The function of inflammation is to eliminate the initial cause of cell injury, clear out necrotic cells and tissues damaged from the original insult and the inflammatory process, and initiate tissue repair. The five classical signs of inflammation are heat, pain, redness, swelling, and loss of function (Latin *calor, dolor, rubor, tumor, and functio laesa*). Inflammation is a generic response, and therefore it is considered as a mechanism of innate immunity, as compared to adaptive immunity, which is specific for each pathogen. Too little inflammation could lead to progressive tissue destruction by the harmful stimulus (*e.g.* bacteria) and compromise the survival of the organism. In contrast, chronic inflammation may lead to a host of diseases, such as hay fever, periodontitis, atherosclerosis, rheumatoid arthritis, and even cancer (*e.g.*, gallbladder carcinoma). Inflammation is therefore normally closely regulated by the body.

Inflammation can be classified as either *acute* or *chronic*. *Acute inflammation* is the initial response of the body to harmful stimuli and is achieved by the increased movement of plasma and leukocytes (especially granulocytes) from the blood into the injured tissues. A series of biochemical events propagates and matures the inflammatory response, involving the local vascular system, the immune system, and various cells within the injured tissue. Prolonged inflammation, known as *chronic inflammation*, leads to a progressive shift in the type of cells present at the site of inflammation, such as mononuclear cells, and is characterized by simultaneous

destruction and healing of the tissue from the inflammatory process. Inflammation is not a synonym for infection. Infection describes the interaction between the action of microbial invasion and the reaction of the body's inflammatory response — the two components are considered together when discussing an infection, and the word is used to imply a microbial invasive cause for the observed inflammatory reaction. Inflammation on the other hand describes purely the body's immunovascular response, whatever the cause may be. But because of how often the two are correlated, words ending in the suffix *-itis* (which refers to inflammation) are sometimes informally described as referring to infection.

For example, the word *urethritis* strictly means only “urethral inflammation”, but clinical health care providers usually discuss urethritis as a urethral infection because urethral microbial invasion is the most common cause of urethritis. It is useful to differentiate inflammation and infection as there are many pathological situations where inflammation is not driven by microbial invasion – for example, atherosclerosis, type III hypersensitivity, trauma, ischaemia. There are also pathological situations where microbial invasion does not result in classic inflammatory response—for example, parasitosis, eosinophilia.

CAUSES

Physical:

- Burns
- Frostbite
- Physical injury, blunt or penetrating
- Foreign bodies, including splinters, dirt and debris
- Trauma
- Ionizing radiation

Biological:

- Infection by pathogens
- Immune reactions due to hypersensitivity
- Stress

Chemical:

- Chemical irritants
- Toxins
- Alcohol

Psychological:

- Excitement

TYPES

• Appendicitis	• Pharyngitis
• Bursitis	• Phlebitis
• Colitis	• Prostatitis
• Cystitis	• RSD/CRPS
• Dermatitis	• Rhinitis
• Encephalitis	• Sinusitis
• Gingivitis	• Tendonitis
• Meningitis	• Testiculitis
• Myelitis	• Tonsillitis
• Nephritis	• Urethritis
• Neuritis	• Vasculitis
• Periodontitis	• Vaginitis

Table. Comparison between acute and chronic inflammation.

	Acute	Chronic
Causative agent	Bacterial pathogens, injured tissues	Persistent acute inflammation due to non-degradable pathogens, viral infection, persistent foreign bodies, or autoimmune reactions
Major cells involved	Neutrophils (primarily), basophils (inflammatory response), and eosinophils (response to helminth worms and parasites), mononuclear cells (monocytes, macrophages)	Mononuclear cells (monocytes, macrophages, lymphocytes, plasma cells), fibroblasts
Primary mediators	Vasoactive amines, eicosanoids	IFN- γ and other cytokines, growth factors, reactive oxygen species, hydrolytic enzymes
Onset	Immediate	Delayed
Duration	Few days	Up to many months, or years
Outcomes	Resolution, abscess formation, chronic inflammation	Tissue destruction, fibrosis, necrosis

Cardinal Signs

Acute inflammation is a short-term process, usually appearing within a few minutes or hours and begins to cease upon the removal of the injurious stimulus. It involves a coordinated and systemic mobilization response locally of various immune, endocrine and neurological mediators of acute inflammation. In a normal healthy response, it becomes activated, clears the pathogen and begins a repair process and then ceases. It is characterized by five cardinal signs:

An acronym that may be used to remember the key symptoms is “PRISH”, for pain, redness, immobility (loss of function), swelling and heat.

The traditional names for signs of inflammation come from Latin:

- Dolor (pain)
- Calor (heat)
- Rubor (redness)
- Tumor (swelling)
- Functio laesa (loss of function)

The first four (classical signs) were described by Celsus (ca. 30 BC–38 AD), while *loss of function* was probably added later by Galen. However, the addition of this fifth sign has also been ascribed to Thomas Sydenham and Virchow. Redness and heat are due to increased blood flow at body core temperature to the inflamed site; swelling is caused by accumulation of fluid; pain is due to the release of chemicals such as bradykinin and histamine that stimulate nerve endings. Loss of function has multiple causes. Acute inflammation of the lung (usually caused in response to pneumonia) does not cause pain unless the inflammation involves the parietal pleura, which does have pain-sensitive nerve endings.

Process of Acute Inflammation

The process of acute inflammation is initiated by resident immune cells already present in the involved tissue, mainly resident macrophages, dendritic cells, histiocytes, Kupffer cells and mast cells. These cells possess surface receptors known as *pattern recognition receptors* (PRRs), which recognize (*i.e.*, bind) two subclasses of molecules: pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs). PAMPs are compounds that are associated with various pathogens, but which are distinguishable from host molecules. DAMPs are compounds that are associated with host-related injury and cell damage.

At the onset of an infection, burn, or other injuries, these cells undergo activation (one of the PRRs recognize a PAMP or DAMP) and release inflammatory mediators responsible for the clinical signs of inflammation. Vasodilation and its resulting increased blood flow causes the redness (*rubor*) and increased heat (*calor*). Increased permeability of the blood vessels results in an exudation (leakage) of plasma proteins and fluid into the tissue (edema), which manifests itself as swelling (*tumor*). Some of the released mediators such as bradykinin increase the sensitivity to pain (hyperalgesia, *dolor*). The mediator molecules also alter the blood vessels to permit the migration of leukocytes, mainly neutrophils and macrophages, outside of the blood vessels (extravasation) into the tissue. The neutrophils migrate along a chemotactic gradient created by the local cells to reach the site of injury. The loss of function (*functio laesa*) is probably the result of a neurological reflex in response to pain. In addition to cell-derived mediators, several acellular biochemical cascade systems consisting of preformed plasma proteins act in parallel to initiate and propagate the inflammatory response. These include the complement system activated by bacteria and the coagulation and fibrinolysis systems activated by necrosis, *e.g.* a burn or a trauma. Acute inflammation may be regarded as the first line of defence against injury. Acute inflammatory response requires constant stimulation to be sustained. Inflammatory mediators are short-lived and are quickly degraded in the tissue. Hence, acute inflammation begins to cease once the stimulus has been removed.

VASCULAR COMPONENT

Vasodilation and Increased Permeability

As defined, acute inflammation is an immunovascular response to an inflammatory stimulus. This means acute inflammation can be broadly divided into a vascular phase that occurs first, followed by a cellular phase involving immune cells (more specifically myeloid granulocytes in the acute setting). The vascular component of acute inflammation involves the movement of plasma fluid, containing important proteins such as fibrin and immunoglobulins (antibodies), into inflamed tissue. Upon contact with PAMPs, tissue macrophages and mastocytes release vasoactive amines such as histamine and serotonin, as well as eicosanoids such as prostaglandin E₂ and leukotriene B₄ to remodel the local vasculature. Macrophages and endothelial cells release nitric oxide. These mediators vasodilate and permeabilize the blood vessels, which results in the net distribution of blood plasma from the vessel into the tissue space. The increased collection of fluid into the tissue causes it to swell (edema). This exuded tissue fluid contains various antimicrobial mediators from the plasma such as complement, lysozyme, antibodies, which can immediately deal damage to microbes, and opsonise the microbes in preparation for the cellular phase. If the inflammatory stimulus is a lacerating wound, exuded platelets, coagulants, plasmin and kinins can clot the wounded area and provide haemostasis in the first instance. These clotting mediators also provide a structural staging framework at the inflammatory tissue site in the form of a fibrin lattice – as would construction scaffolding at a construction site – for the purpose of aiding phagocytic debridement and wound repair later on. Some of the exuded tissue fluid is also funnelled by lymphatics to the regional lymph nodes, flushing bacteria along to start the recognition and attack phase of the adaptive immune system.

Acute inflammation is characterized by marked vascular changes, including vasodilation, increased permeability and increased blood flow, which are induced by the actions of various inflammatory mediators. Vasodilation occurs first at the arteriole level, progressing to the capillary level, and brings about a net increase in the amount of blood present, causing the redness and heat of inflammation. Increased permeability of the vessels results in the movement of plasma into the tissues, with resultant *stasis* due to the increase in the concentration of the cells within blood – a condition characterized by enlarged vessels packed with cells. Stasis allows leukocytes to marginate (move) along the endothelium, a process critical to their recruitment into the tissues. Normal flowing blood prevents this, as the shearing force along the periphery of the vessels moves cells in the blood into the middle of the vessel.

Plasma Cascade Systems

- The complement system, when activated, creates a cascade of chemical reactions that promotes opsonization, chemotaxis, and agglutination, and produces the MAC.
- The kinin system generates proteins capable of sustaining vasodilation and other physical inflammatory effects.
- The coagulation system or *clotting cascade*, which forms a protective protein mesh over sites of injury.
- The fibrinolysis system, which acts in opposition to the *coagulation system*, to counterbalance clotting and generate several other inflammatory mediators.

Plasma-derived Mediators

Name	Produced by	Description
Bradykinin	Kinin system	A vasoactive protein that is able to induce vasodilation, increase vascular permeability, cause smooth muscle contraction, and induce pain.
C3	Complement system	Cleaves to produce C3a and C3b. C3a stimulates histamine release by mast cells, thereby producing vasodilation. C3b is able to bind to bacterial cell walls and act as an opsonin, which marks the invader as a target for phagocytosis.
C5a	Complement system	Stimulates histamine release by mast cells, thereby producing vasodilation. It is also able to act as a chemoattractant to direct cells via chemotaxis to the site of inflammation.
Factor XII (Hageman Factor)	Liver	A protein that circulates inactive, until activated by collagen, platelets, or exposed basement membranes via conformational change. When activated, it in turn is able to activate three plasma systems involved in inflammation: the kinin system, fibrinolysis system, and coagulation system.
Membrane attack complex	Complement system	A complex of the complement proteins C5b, C6, C7, C8, and multiple units of C9. The combination and activation of this range of complement proteins forms the membrane attack complex, which is able to insert into bacterial cell walls and causes cell lysis with ensuing bacterial death.
Plasmin	Fibrinolysis system	Able to break down fibrin clots, cleave complement protein C3, and activate Factor XII.
Thrombin	Coagulation system	Cleaves the soluble plasma protein fibrinogen to produce insoluble fibrin, which aggregates to form a blood clot. Thrombin can also bind to cells via the PAR1 receptor to trigger several other inflammatory responses, such as production of chemokines and nitric oxide.

CELLULAR COMPONENT

The *cellular component* involves leukocytes, which normally reside in blood and must move into the inflamed tissue via *extravasation* to aid in inflammation. Some act as phagocytes, ingesting bacteria, viruses, and cellular debris. Others release enzymatic granules that damage pathogenic invaders. Leukocytes also release inflammatory mediators that develop and maintain the inflammatory response. In general, acute inflammation is mediated by granulocytes, whereas chronic inflammation is mediated by mononuclear cells such as monocytes and lymphocytes.

Leukocyte Extravasation

Various leukocytes, particularly neutrophils, are critically involved in the initiation and maintenance of inflammation. These cells must be able to move to the site of injury from their usual location in the blood, therefore mechanisms exist to recruit and direct leukocytes to the appropriate place. The process of leukocyte movement from the blood to the tissues through the blood vessels is known as *extravasation*, and can be broadly divided up into a number of steps:

- *Leukocyte margination and endothelial adhesion:* The white blood cells within the vessels which are generally centrally located move peripherally towards the walls of the vessels. Activated macrophages in the tissue release cytokines such as IL-1 and TNF α , which in turn leads to production of chemokines that bind to proteoglycans forming gradient in the inflamed tissue and along the endothelial wall. Inflammatory cytokines induce the immediate expression of P-selectin on endothelial cell surfaces and P-selectin binds weakly to carbohydrate ligands on the surface of leukocytes and causes them to “roll” along the endothelial surface as bonds are made and broken. Cytokines released from injured cells induce the expression of E-selectin on endothelial cells, which functions similarly to P-selectin. Cytokines also induce the expression of integrin ligands such as ICAM-1 and VCAM-1 on endothelial cells, which mediate the adhesion and further slow leukocytes down. These weakly bound leukocytes are free to detach if not activated by chemokines produced in injured tissue after signal transduction via respective G protein-coupled receptors that activates integrins on the leukocyte surface for firm adhesion. Such activation increases the affinity of bound integrin receptors for ICAM-1 and VCAM-1 on the endothelial cell surface, firmly binding the leukocytes to the endothelium.
- Migration across the endothelium, known as *transmigration*, via the process of diapedesis: Chemokine gradients stimulate the adhered leukocytes to move between adjacent endothelial cells. The endothelial cells retract and the leukocytes pass through the basement membrane into the surrounding tissue using adhesion molecules such as ICAM-1.
- *Movement of leukocytes within the tissue via chemotaxis:* Leukocytes reaching the tissue interstitium bind to extracellular matrix proteins via expressed integrins and CD44 to prevent them from leaving the site. A variety of molecules behave as chemoattractants, for example, C3a or C5, and cause the leukocytes to move along a chemotactic gradient towards the source of inflammation.

Phagocytosis

Extravasated neutrophils in the cellular phase come into contact with microbes at the inflamed tissue. Phagocytes express cell-surface endocytic pattern recognition receptors (PRRs) that have affinity and efficacy against non-specific microbe-associated molecular patterns (PAMPs). Most PAMPs that bind to endocytic PRRs and initiate phagocytosis are cell wall components, including complex carbohydrates such as mannans and β -glucans, lipopolysaccharides (LPS), peptidoglycans, and surface proteins. Endocytic PRRs on phagocytes reflect these molecular patterns, with C-type lectin receptors binding to mannans and β -glucans, and scavenger receptors

binding to LPS. Upon endocytic PRR binding, actin-myosin cytoskeletal rearrangement adjacent to the plasma membrane occurs in a way that endocytoses the plasma membrane containing the PRR-PAMP complex, and the microbe. Phosphatidylinositol and Vps34-Vps15-Beclin1 signalling pathways have been implicated to traffic the endocytosed phagosome to intracellular lysosomes, where fusion of the phagosome and the lysosome produces a phagolysosome. The reactive oxygen species, superoxides and hypochlorite bleach within the phagolysosomes then kill microbes inside the phagocyte. Phagocytic efficacy can be enhanced by opsonization. Plasma derived complement C3b and antibodies that exude into the inflamed tissue during the vascular phase bind to and coat the microbial antigens. As well as endocytic PRRs, phagocytes also express opsonin receptors Fc receptor and complement receptor 1 (CR1), which bind to antibodies and C3b, respectively. The co-stimulation of endocytic PRR and opsonin receptor increases the efficacy of the phagocytic process, enhancing the lysosomal elimination of the infective agent.

Cell-derived Mediators

Name	Type	Source	Description
Lysosome granules	Enzymes	Granulocytes	These cells contain a large variety of enzymes that perform a number of functions. Granules can be classified as either specific or azurophilic depending upon the contents, and are able to break down a number of substances, some of which may be plasma-derived proteins that allow these enzymes to act as inflammatory mediators.
Histamine	Monoamine	Mast cells and basophils	Stored in preformed granules, histamine is released in response to a number of stimuli. It causes arteriole dilation, increased venous permeability, and a wide variety of organ-specific effects.
IFN- γ	Cytokine	T-cells, NK cells	Antiviral, immunoregulatory, and anti-tumour properties. This interferon was originally called macrophage-activating factor, and is especially important in the maintenance of chronic inflammation.
IL-8	Chemokine	Primarily macrophages	Activation and chemoattraction of neutrophils, with a weak effect on monocytes and eosinophils.
Leukotriene B4	Eicosanoid	Leukocytes, cancer cells	Able to mediate leukocyte adhesion and activation, allowing them to bind to the endothelium and migrate across it. In neutrophils, it is also a potent chemoattractant, and is able to induce the formation of reactive oxygen species and the release of lysosomal enzymes by these cells.
LTC4, LTD4	Eicosanoid	eosinophils, mast cells, macrophages	These three Cysteine-containing leukotrienes contract lung airways, increase micro-vascular permeability, stimulate mucus secretion, and promote eosinophil-based inflammation in the lung, skin, nose, eye, and other tissues.
5-oxo-eicosatetraenoic acid	Eicosanoid	leukocytes, cancer cells	Potent stimulator of neutrophil chemotaxis, lysosome enzyme release, and reactive oxygen species formation; monocyte chemotaxis; and with even greater potency eosinophil chemotaxis, lysosome enzyme release, and reactive oxygen species formation.
5-HETE	Eicosanoid	Leukocytes	Metabolic precursor to 5-Oxo-eicosatetraenoic acid, it is a less potent stimulator of neutrophil chemotaxis, lysosome enzyme release, and reactive oxygen species formation; monocyte chemotaxis; and eosinophil chemotaxis, lysosome enzyme release, and reactive oxygen species formation.
Prostaglandins	Eicosanoid	Mast cells	A group of lipids that can cause vasodilation, fever, and pain.
Nitric oxide	Soluble gas	Macrophages, endothelial cells, some neurons	Potent vasodilator, relaxes smooth muscle, reduces platelet aggregation, aids in leukocyte recruitment, direct antimicrobial activity in high concentrations.
TNF- α and IL-1	Cytokines	Primarily macrophages	Both affect a wide variety of cells to induce many similar inflammatory reactions: fever, production of cytokines, endothelial gene regulation, chemotaxis, leukocyte adherence, activation of fibroblasts. Responsible for the systemic effects of inflammation, such as loss of appetite and increased heart rate. TNF- α inhibits osteoblast differentiation.
Tryptase	Enzymes	Mast Cells	This serine protease is believed to be exclusively stored in mast cells and secreted, along with histamine, during mast cell activation.

MORPHOLOGIC PATTERNS

Specific patterns of acute and chronic inflammation are seen during particular situations that arise in the body, such as when inflammation occurs on an epithelial surface, or pyogenic bacteria are involved.

- *Granulomatous inflammation:* Characterised by the formation of granulomas, they are the result of a limited but diverse number of diseases, which include among others tuberculosis, leprosy, sarcoidosis, and syphilis.
- *Fibrinous inflammation:* Inflammation resulting in a large increase in vascular permeability allows fibrin to pass through the blood vessels. If an appropriate *procoagulative* stimulus is present, such as cancer cells, a fibrinous exudate is deposited. This is commonly seen in serous cavities, where the conversion of fibrinous exudate into a scar can occur between serous membranes, limiting their function. The deposit sometimes forms a pseudomembrane sheet. During inflammation of the intestine (Pseudomembranous colitis), pseudomembranous tubes can be formed.
- *Purulent inflammation:* Inflammation resulting in large amount of pus, which consists of neutrophils, dead cells, and fluid. Infection by pyogenic bacteria such as staphylococci is characteristic of this kind of inflammation. Large, localised collections of pus enclosed by surrounding tissues are called abscesses.
- *Serous inflammation:* Characterised by the copious effusion of non-viscous serous fluid, commonly produced by mesothelial cells of serous membranes, but may be derived from blood plasma. Skin blisters exemplify this pattern of inflammation.
- *Ulcerative inflammation:* Inflammation occurring near an epithelium can result in the necrotic loss of tissue from the surface, exposing lower layers. The subsequent excavation in the epithelium is known as an ulcer.

INFLAMMATORY DISORDERS

Inflammatory abnormalities are a large group of disorders that underlie a vast variety of human diseases. The immune system is often involved with inflammatory disorders, demonstrated in both allergic reactions and some myopathies, with many immune system disorders resulting in abnormal inflammation. Non-immune diseases with causal origins in inflammatory processes include cancer, atherosclerosis, and ischemic heart disease.

Examples of disorders associated with inflammation include:

- Acne vulgaris
- Asthma
- Autoimmune diseases
- Autoinflammatory diseases
- Celiac disease
- Chronic prostatitis
- Colitis
- Diverticulitis
- Glomerulonephritis
- Hidradenitis suppurativa
- Hypersensitivities
- Inflammatory bowel diseases
- Interstitial cystitis
- Lichen planus
- Mast Cell Activation Syndrome
- Mastocytosis
- Otitis

- Pelvic inflammatory disease
- Reperfusion injury
- Rheumatic fever
- Rheumatoid arthritis
- Rhinitis
- Sarcoidosis
- Transplant rejection
- Vasculitis

Atherosclerosis

Atherosclerosis, formerly considered a bland lipid storage disease, actually involves an ongoing inflammatory response. Recent advances in basic science have established a fundamental role for inflammation in mediating all stages of this disease from initiation through progression and, ultimately, the thrombotic complications of atherosclerosis. These new findings provide important links between risk factors and the mechanisms of atherogenesis. Clinical studies have shown that this emerging biology of inflammation in atherosclerosis applies directly to human patients. Elevation in markers of inflammation predicts outcomes of patients with acute coronary syndromes, independently of myocardial damage. In addition, low-grade chronic inflammation, as indicated by levels of the inflammatory marker C-reactive protein, prospectively defines risk of atherosclerotic complications, thus adding to prognostic information provided by traditional risk factors. Moreover, certain treatments that reduce coronary risk also limit inflammation. In the case of lipid lowering with statins, this anti-inflammatory effect does not appear to correlate with reduction in low-density lipoprotein levels. These new insights into inflammation in atherosclerosis not only increase our understanding of this disease but also have practical clinical applications in risk stratification and targeting of therapy for this scourge of growing worldwide importance.

Allergy

An allergic reaction, formally known as type 1 hypersensitivity, is the result of an inappropriate immune response triggering inflammation, vasodilation, and nerve irritation. A common example is hay fever, which is caused by a hypersensitive response by mast cells to allergens. Pre-sensitised mast cells respond by degranulating, releasing vasoactive chemicals such as histamine. These chemicals propagate an excessive inflammatory response characterised by blood vessel dilation, production of pro-inflammatory molecules, cytokine release, and recruitment of leukocytes. Severe inflammatory response may mature into a systemic response known as anaphylaxis.

Myopathies

Inflammatory myopathies are caused by the immune system inappropriately attacking components of muscle, leading to signs of muscle inflammation. They may occur in conjunction with other immune disorders, such as systemic sclerosis, and include dermatomyositis, polymyositis, and inclusion body myositis.

Leukocyte Defects

Due to the central role of leukocytes in the development and propagation of inflammation, defects in leukocyte functionality often result in a decreased capacity for inflammatory defence with subsequent vulnerability to infection. Dysfunctional leukocytes may be unable to correctly bind to blood vessels due to surface receptor mutations, digest bacteria (Chédiak–Higashi syndrome), or produce microbicides (chronic granulomatous disease). In addition, diseases affecting the bone marrow may result in abnormal or few leukocytes.

Pharmacological

Certain drugs or exogenous chemical compounds are known to affect inflammation. Vitamin A deficiency causes an increase in inflammatory responses, and anti-inflammatory drugs work specifically by inhibiting the enzymes that produce inflammatory eicosanoids. Certain illicit drugs such as cocaine and ecstasy may exert some of their detrimental effects by activating transcription factors intimately involved with inflammation (e.g. NF- κ B).

Cancer

Inflammation orchestrates the microenvironment around tumours, contributing to proliferation, survival and migration. Cancer cells use selectins, chemokines and their receptors for invasion, migration and metastasis. On the other hand, many cells of the immune system contribute to cancer immunology, suppressing cancer. Molecular intersection between receptors of steroid hormones, which have important effects on cellular development, and transcription factors that play key roles in inflammation, such as NF- κ B, may mediate some of the most critical effects of inflammatory stimuli on cancer cells. This capacity of a mediator of inflammation to influence the effects of steroid hormones in cells, is very likely to affect carcinogenesis on the one hand; on the other hand, due to the modular nature of many steroid hormone receptors, this interaction may offer ways to interfere with cancer progression, through targeting of a specific protein domain in a specific cell type. Such an approach may limit side effects that are unrelated to the tumor of interest, and may help preserve vital homeostatic functions and developmental processes in the organism. According to a review of 2009, recent data suggests that cancer-related inflammation (CRI) may lead to accumulation of random genetic alterations in cancer cells.

Importance of Inflammation in Cancer

In 1863, Rudolf Virchow hypothesized that the origin of cancer was at sites of chronic inflammation. At present, chronic inflammation is estimated to contribute to approximately 15 per cent to 25 per cent of human cancers.

Mediators and DNA Damage in Cancer

An inflammatory mediator is a messenger that acts on blood vessels and/or cells to promote an inflammatory response. Inflammatory mediators that contribute to neoplasia include prostaglandins, inflammatory cytokines such as IL-1 β , TNF- α , IL-6 and IL-15 and chemokines such as IL-8 and GRO- α . These inflammatory mediators, and others, orchestrate an environment that fosters proliferation and survival. Inflammation also causes DNA damages due to the induction of reactive oxygen species (ROS) by various intracellular inflammatory mediators. In addition, leukocytes and other phagocytic cells attracted to the site of inflammation induce DNA damages in proliferating cells through their generation of ROS and reactive nitrogen species (RNS). ROS and RNS are normally produced by these cells to fight infection. ROS, alone, cause more than 20 types of DNA damage. Oxidative DNA damages cause both mutations and epigenetic alterations. RNS also cause mutagenic DNA damages. A normal cell may undergo carcinogenesis to become a cancer cell if it is frequently subjected to DNA damage during long periods of chronic inflammation. DNA damages may cause genetic mutations due to inaccurate repair. In addition, mistakes in the DNA repair process may cause epigenetic alterations. Mutations and epigenetic alterations that are replicated and provide a selective advantage during somatic cell proliferation may be carcinogenic. Genome-wide analyses of human cancer tissues reveal that a single typical cancer cell may possess roughly 100 mutations in coding regions, 10-20 of which are “driver mutations” that contribute to cancer development. However, chronic inflammation also causes epigenetic changes such as DNA methylations, that are often more common than mutations. Typically, several hundreds to thousands of genes are methylated in a cancer cell. Sites of oxidative damage in chromatin can recruit complexes that contain DNA

methyltransferases (DNMTs), a histone deacetylase (SIRT1), and a histone methyltransferase (EZH2), and thus induce DNA methylation. DNA methylation of a CpG island in a promoter region may cause silencing of its downstream gene. DNA repair genes, in particular, are frequently inactivated by methylation in various cancers. A 2018 report evaluated the relative importance of mutations and epigenetic alterations in progression to two different types of cancer. This report showed that epigenetic alterations were much more important than mutations in generating gastric cancers (associated with inflammation). However, mutations and epigenetic alterations were of roughly equal importance in generating esophageal squamous cell cancers (associated with tobacco chemicals and acetaldehyde, a product of alcohol metabolism).

HIV and AIDS

It has long been recognized that infection with HIV is characterized not only by development of profound immunodeficiency but also by sustained inflammation and immune activation. A substantial body of evidence implicates chronic inflammation as a critical driver of immune dysfunction, premature appearance of aging-related diseases, and immune deficiency. Many now regard HIV infection not only as an evolving virus-induced immunodeficiency but also as chronic inflammatory disease. Even after the introduction of effective antiretroviral therapy (ART) and effective suppression of viremia in HIV-infected individuals, chronic inflammation persists. Animal studies also support the relationship between immune activation and progressive cellular immune deficiency: SIVsm infection of its natural nonhuman primate hosts, the sooty mangabey, causes high-level viral replication but limited evidence of disease. This lack of pathogenicity is accompanied by a lack of inflammation, immune activation and cellular proliferation. In sharp contrast, experimental SIVsm infection of rhesus macaque produces immune activation and AIDS-like disease with many parallels to human HIV infection.

Delineating how CD4 T cells are depleted and how chronic inflammation and immune activation are induced lies at the heart of understanding HIV pathogenesis—one of the top priorities for HIV research by the Office of AIDS Research, National Institutes of Health. Recent studies demonstrated that caspase-1-mediated pyroptosis, a highly inflammatory form of programmed cell death, drives CD4 T-cell depletion and inflammation by HIV. These are the two signature events that propel HIV disease progression to AIDS. Pyroptosis appears to create a pathogenic vicious cycle in which dying CD4 T cells and other immune cells (including macrophages and neutrophils) release inflammatory signals that recruit more cells into the infected lymphoid tissues to die. The feed-forward nature of this inflammatory response produces chronic inflammation and tissue injury. Identifying pyroptosis as the predominant mechanism that causes CD4 T-cell depletion and chronic inflammation, provides novel therapeutic opportunities, namely caspase-1 which controls the pyroptotic pathway. In this regard, pyroptosis of CD4 T cells and secretion of pro-inflammatory cytokines such as IL-1 β and IL-18 can be blocked in HIV-infected human lymphoid tissues by addition of the caspase-1 inhibitor VX-765, which has already proven to be safe and well tolerated in phase II human clinical trials. These findings could propel development of an entirely new class of “anti-AIDS” therapies that act by targeting the host rather than the virus. Such agents would almost certainly be used in combination with ART. By promoting “tolerance” of the virus instead of suppressing its replication, VX-765 or related drugs may mimic the evolutionary solutions occurring in multiple monkey hosts (*e.g.* the sooty mangabey) infected with species-specific lentiviruses that have led to a lack of disease, no decline in CD4 T-cell counts, and no chronic inflammation.

Resolution of Inflammation

The inflammatory response must be actively terminated when no longer needed to prevent unnecessary “bystander” damage to tissues. Failure to do so results in chronic inflammation, and cellular destruction. Resolution of inflammation occurs by different mechanisms in different tissues. Mechanisms that serve to terminate inflammation include:

- Short half-life of inflammatory mediators *in vivo*.
- Production and release of transforming growth factor (TGF) beta from macrophages
- Production and release of interleukin 10 (IL-10)
- Production of anti-inflammatory specialized proresolving mediators, *i.e.* lipoxins, resolvins, maresins, and neuroprotectins
- Downregulation of pro-inflammatory molecules, such as leukotrienes.
- Upregulation of anti-inflammatory molecules such as the interleukin 1 receptor antagonist or the soluble tumor necrosis factor receptor (TNFR)
- Apoptosis of pro-inflammatory cells
- Desensitization of receptors.
- Increased survival of cells in regions of inflammation due to their interaction with the extracellular matrix (ECM)
- Downregulation of receptor activity by high concentrations of ligands
- Cleavage of chemokines by matrix metalloproteinases (MMPs) might lead to production of anti-inflammatory factors.

Connection to Depression

There is evidence for a link between inflammation and depression. Inflammatory processes can be triggered by negative cognitions or their consequences, such as stress, violence, or deprivation. Thus, negative cognitions can cause inflammation that can, in turn, lead to depression. In addition there is increasing evidence that inflammation can cause depression because of the increase of cytokines, setting the brain into a “sickness mode”. Classical symptoms of being physically sick like lethargy show a large overlap in behaviors that characterize depression. Levels of cytokines tend to increase sharply during depressive episodes in manics and drop off during remission. Furthermore, it has been shown in clinical trials that anti-inflammatory medicines taken in addition to antidepressants not only significantly improves symptoms but also increases the proportion of subjects positively responding to treatment. Inflammations that lead to serious depression could be caused by common infections such as those caused by a virus, bacteria or even parasites.

SYSTEMIC EFFECTS

An infectious organism can escape the confines of the immediate tissue via the circulatory system or lymphatic system, where it may spread to other parts of the body. If an organism is not contained by the actions of acute inflammation it may gain access to the lymphatic system via nearby lymph vessels. An infection of the lymph vessels is known as lymphangitis, and infection of a lymph node is known as lymphadenitis. When lymph nodes cannot destroy all pathogens, the infection spreads further. A pathogen can gain access to the bloodstream through lymphatic drainage into the circulatory system. When inflammation overwhelms the host, systemic inflammatory response syndrome is diagnosed. When it is due to infection, the term sepsis is applied, with the terms bacteremia being applied specifically for bacterial sepsis and viremia specifically to viral sepsis. Vasodilation and organ dysfunction are serious problems associated with widespread infection that may lead to septic shock and death.

Acute-phase Proteins

Inflammation also induces high systemic levels of acute-phase proteins. In acute inflammation, these proteins prove beneficial; however, in chronic inflammation they can contribute to amyloidosis. These proteins include C-reactive protein, serum amyloid A, and serum amyloid P, which cause a range of systemic effects including:

- Fever
- Increased blood pressure
- Decreased sweating
- Malaise
- Loss of appetite
- Somnolence

Leukocyte Numbers

Inflammation often affects the numbers of leukocytes present in the body:

- Leukocytosis is often seen during inflammation induced by infection, where it results in a large increase in the amount of leukocytes in the blood, especially immature cells. Leukocyte numbers usually increase to between 15 000 and 20 000 cells per microliter, but extreme cases can see it approach 100 000 cells per microliter. Bacterial infection usually results in an increase of neutrophils, creating neutrophilia, whereas diseases such as asthma, hay fever, and parasite infestation result in an increase in eosinophils, creating eosinophilia.
- Leukopenia can be induced by certain infections and diseases, including viral infection, *Rickettsia* infection, some protozoa, tuberculosis, and some cancers.

Systemic Inflammation and Obesity

With the discovery of interleukins (IL), the concept of systemic inflammation developed. Although the processes involved are identical to tissue inflammation, systemic inflammation is not confined to a particular tissue but involves the endothelium and other organ systems.

Chronic inflammation is widely observed in obesity. The obese commonly have many elevated markers of inflammation, including:

- IL-6 (Interleukin-6)
- IL-8 (Interleukin-8)
- IL-18 (Interleukin-18)
- TNF- α (Tumor necrosis factor-alpha)
- CRP (C-reactive protein)
- Insulin
- Blood glucose
- Leptin

Low-grade chronic inflammation is characterized by a two- to threefold increase in the systemic concentrations of cytokines such as TNF- α , IL-6, and CRP. Waist circumference correlates significantly with systemic inflammatory response. A predominant factor in this correlation is due to the autoimmune response triggered by adiposity, whereby immune cells may mistake fatty deposits for intruders. The body attacks fat like it does bacteria and fungi. When expanded fat cells leak or break open, macrophages mobilize to clean up and embed into the adipose tissue. Then macrophages release inflammatory chemicals, including TNF- α and IL-6. TNF's primary role is to regulate the immune cells and induce inflammation. White blood cells then assist by releasing more cytokines. This link between adiposity and inflammation has been shown to produce 10–35 per cent of IL-6 in a resting individual, and this production increases with increasing adiposity. Loss of white adipose tissue reduces levels of inflammation markers. The association of systemic inflammation with insulin resistance and atherosclerosis is the subject of intense research. In the obese mouse models, inflammation and macrophage-specific genes are upregulated in white adipose tissue (WAT). There were also signs of dramatic increase in circulating insulin level, adipocyte lipolysis and formation of multinucleate giant cells. The fat-derived protein

called angiopoietin-like protein 2 (Angptl2) elevates in fat tissues. Higher than normal Angptl2 level in fat tissues develop inflammation as well as insulin and leptin resistance. Stored fat secretes leptin to signal satiety. Leptin resistance plays a role in the process where appetite overrules the message of satiety. Angptl2 then starts an inflammatory cascade causing blood vessels to remodel and attract macrophages. Angptl2 is an adipocyte-derived inflammatory mediator linking obesity to systemic insulin resistance. It is possible that, as an inflammatory marker, leptin responds specifically to adipose-derived inflammatory cytokines. C-reactive protein (CRP) is generated at a higher level in obese people. It raises when there is inflammation throughout the body. Mild elevation in CRP increases risk of heart attacks, strokes, high blood pressure, muscle weakness and fragility.

Systemic Inflammation and Overeating

Hyperglycemia induces IL-6 production from endothelial cells and macrophages. Meals high in saturated fat, as well as meals high in calories have been associated with increases in inflammatory markers. In addition, interstitial abdominal adiposity (also referred to as accumulated intra-abdominal fat) may be a factor in increasing systemic risk for multiple inflammatory diseases. Although the exact mechanisms are still being investigated, a review published in 2010 suggested that significant growth of adipose tissue in response to overeating can evoke a chronic inflammatory response.

OUTCOMES

The outcome in a particular circumstance will be determined by the tissue in which the injury has occurred and the injurious agent that is causing it. Here are the possible outcomes to inflammation:

- Resolution
- The complete restoration of the inflamed tissue back to a normal status. Inflammatory measures such as vasodilation, chemical production, and leukocyte infiltration cease, and damaged parenchymal cells regenerate. In situations where limited or short-lived inflammation has occurred this is usually the outcome.
- Fibrosis
- Large amounts of tissue destruction, or damage in tissues unable to regenerate, cannot be regenerated completely by the body. Fibrous scarring occurs in these areas of damage, forming a scar composed primarily of collagen. The scar will not contain any specialized structures, such as parenchymal cells, hence functional impairment may occur.
- Abscess formation
- A cavity is formed containing pus, an opaque liquid containing dead white blood cells and bacteria with general debris from destroyed cells.
- Chronic inflammation
- In acute inflammation, if the injurious agent persists then chronic inflammation will ensue. This process, marked by inflammation lasting many days, months or even years, may lead to the formation of a chronic wound. Chronic inflammation is characterised by the dominating presence of macrophages in the injured tissue. These cells are powerful defensive agents of the body, but the toxins they release (including reactive oxygen species) are injurious to the organism's own tissues as well as invading agents. As a consequence, chronic inflammation is almost always accompanied by tissue destruction.

DIET AND INFLAMMATION

The Dietary Inflammatory Index (DII) is a score (number) that describes the potential of diet to modulate systemic inflammation within the body. As stated chronic inflammation is linked to most chronic diseases including arthritis, many types of cancer, cardiovascular diseases, inflammatory bowel diseases, and diabetes.

EXERCISE AND INFLAMMATION

Exercise-induced Acute Inflammation

Acute inflammation of the muscle cells, as understood in exercise physiology, can result after induced eccentric and concentric muscle training. Participation in eccentric training and conditioning, including resistance training and activities that emphasize eccentric lengthening of the muscle including downhill running on a moderate to high incline can result in considerable soreness within 24 to 48 hours, even though blood lactate levels, previously thought to cause muscle soreness, were much higher with level running. This delayed onset muscle soreness (DOMS) results from structural damage to the contractile filaments and z-disks, which has been noted especially in marathon runners whose muscle fibers revealed remarkable damage to the muscle fibers after both training and marathon competition. The onset and timing of this gradient damage to the muscle parallels the degree of muscle soreness experienced by the runners.

Z-disks are the point of contact for the contractile proteins. They provide structural support for transmission of force when muscle fibers are activated to shorten. However, in marathon runners and those who subscribe to the overload principle to enhance their muscles, show moderate Z-disk streaming and major disruption of thick and thin filaments in parallel groups of sarcomeres as a result of the force of eccentric actions or stretching of tightened muscle fibers. This disruption of muscle fibers triggers white blood cells to increase following induced muscle soreness, leading to the inflammatory response observation from induced muscle soreness. Elevations in plasma enzymes, myoglobinemia, and abnormal muscle histology and ultrastructure are concluded to be associated with inflammatory response. High tension in the contractile-elastic system of muscle results in structural damage to the muscle fibre and plasmalemma and its epimysium, perimysium, and/or endomysium. The myosium damage disrupts calcium homeostasis in injured fibers and fibre bundles, resulting in necrosis that peaks about 48 hours after exercise. The products of macrophage activity and intracellular contents (such as histamines, kinins, and K) accumulate outside cells. These substances then stimulate free nerve endings in the muscle; a process that appears accentuated by eccentric exercise, in which large forces are distributed over a relatively small cross-sectional area of the muscle.

Post-inflammatory Muscle Growth and Repair

There is a known relationship between inflammation and muscle growth. For instance, high doses of anti-inflammatory medicines (*e.g.*, NSAIDs) are able to blunt muscle growth. Cold therapy has been shown to negatively affect muscle growth as well. Reducing inflammation results in decreased macrophage activity and lower levels of IGF-1. Acute effects of cold therapy on training adaptations show reduced satellite cell proliferation. Long term effects include less muscular hypertrophy and an altered cell structure of muscle fibers. It has been further theorized that the acute localized inflammatory responses to muscular contraction during exercise, as described above, are a necessary precursor to muscle growth. As a response to muscular contractions, the acute inflammatory response initiates the breakdown and removal of damaged muscle tissue. Muscles can synthesize cytokines in response to contractions, such that the cytokines interleukin-1 beta (IL-1 β), TNF- α , and IL-6 are expressed in skeletal muscle up to 5 days after exercise. In particular, the increase in levels of IL-6 (interleukin 6), a myokine, can reach up to one hundred times that of resting levels. Depending on volume, intensity, and other training factors, the IL-6 increase associated with training initiates about 4 hours after resistance training and remains elevated for up to 24 hours. These acute increases in cytokines, as a response to muscle contractions, help initiate the process of muscle repair and growth by activating satellite cells within the inflamed muscle. Satellite cells are crucial for skeletal muscle adaptation to exercise. They contribute to hypertrophy by providing new myonuclei and repair damaged segments of mature myofibers for successful regeneration following injury-

or exercise-induced muscle damage; high-level powerlifters can have up to 100 per cent more satellite cells than untrained controls. A rapid and transient localization of the IL-6 receptor and increased IL-6 expression occurs in satellite cells following contractions. IL-6 has been shown to mediate hypertrophic muscle growth both *in vitro* and *in vivo*. Unaccustomed exercise can increase IL-6 by up to sixfold at 5 hours post-exercise and threefold 8 days after exercise. Also telling is the fact that NSAIDs can decrease satellite cell response to exercise, thereby reducing exercise-induced protein synthesis. The increase in cytokines (myokines) after resistance exercise coincides with the decrease in levels of myostatin, a protein that inhibits muscle differentiation and growth. The cytokine response to resistance exercise and moderate-intensity running occur differently, with the latter causing a more prolonged response, especially at the 12–24 hour mark. Developing research has demonstrated that many of the benefits of exercise are mediated through the role of skeletal muscle as an endocrine organ. That is, contracting muscles release multiple substances known as myokines, including but not limited to those cited in the above description, which promote the growth of new tissue, tissue repair, and various anti-inflammatory functions, which in turn reduce the risk of developing various inflammatory diseases. The new view that muscle is an endocrine organ is transforming our understanding of exercise physiology and with it, of the role of inflammation in adaptation to stress.

Chronic Inflammation and Muscle Loss

Both chronic and extreme inflammation are associated with disruptions of anabolic signals initiating muscle growth. Chronic inflammation has been implicated as part of the cause of the muscle loss that occurs with aging. Increased protein levels of myostatin have been described in patients with diseases characterized by chronic low-grade inflammation. Increased levels of TNF- α can suppress the AKT/mTOR pathway, a crucial pathway for regulating skeletal muscle hypertrophy, thereby increasing muscle catabolism. Cytokines may antagonize the anabolic effects of insulin-like growth factor 1 (IGF-1). In the case of sepsis, an extreme whole body inflammatory state, the synthesis of both myofibrillar and sarcoplasmic proteins are inhibited, with the inhibition taking place preferentially in fast-twitch muscle fibers. Sepsis is also able to prevent leucine from stimulating muscle protein synthesis. In animal models, when inflammation is created, mTOR loses its ability to be stimulated by muscle growth.

Exercise as a Treatment for Inflammation

Regular physical activity is reported to decrease markers of inflammation, although the correlation is imperfect and seems to reveal differing results contingent upon training intensity. For instance, while baseline measurements of circulating inflammatory markers do not seem to differ greatly between healthy trained and untrained adults, long-term training may help reduce chronic low-grade inflammation. On the other hand, levels of the anti-inflammatory myokine IL-6 (interleukin 6) remained elevated longer into the recovery period following an acute bout of exercise in patients with inflammatory diseases, relative to the recovery of healthy controls. It may well be that low-intensity training can reduce resting pro-inflammatory markers (CRP, IL-6), while moderate-intensity training has milder and less-established anti-inflammatory benefits. There is a strong relationship between exhaustive exercise and chronic low-grade inflammation. Marathon running may enhance IL-6 levels as much as 100 times over normal and increases total leucocyte count and neutrophil mobilization.

Regarding the above, IL-6 had previously been classified as a proinflammatory cytokine. Therefore, it was first thought that the exercise-induced IL-6 response was related to muscle damage. However, it has become evident that eccentric exercise is not associated with a larger increase in plasma IL-6 than exercise involving concentric “nondamaging” muscle contractions. This finding clearly demonstrates that muscle damage is not required to provoke an increase in plasma IL-6 during exercise. As a matter of fact, eccentric exercise may result in a delayed peak and a much slower decrease of plasma IL-6 during recovery.

Recent work has shown that both upstream and downstream signalling pathways for IL-6 differ markedly between myocytes and macrophages. It appears that unlike IL-6 signalling in macrophages, which is dependent upon activation of the NFκB signalling pathway, intramuscular IL-6 expression is regulated by a network of signalling cascades, including the Ca²⁺/NFAT and glycogen/p38 MAPK pathways. Thus, when IL-6 is signalling in monocytes or macrophages, it creates a pro-inflammatory response, whereas IL-6 activation and signalling in muscle is totally independent of a preceding TNF-response or NFκB activation, and is anti-inflammatory. Several studies show that markers of inflammation are reduced following longer-term behavioural changes involving both reduced energy intake and a regular programme of increased physical activity, and that, in particular, IL-6 was miscast as an inflammatory marker. For example, the anti-inflammatory effects of IL-6 have been demonstrated by IL-6 stimulating the production of the classical anti-inflammatory cytokines IL-1ra and IL-10.

As such, individuals pursuing exercise as a means to treat the causal factors underlying chronic inflammation are pursuing a course of action strongly supported by current research, as an inactive lifestyle is strongly associated with the development and progression of multiple inflammatory diseases. Note that cautions regarding over-exertion may apply in certain cases, as discussed above, though this concern rarely applies to the general population.

Signal-to-noise Theory

Given that localized acute inflammation is a necessary component for muscle growth, and that chronic low-grade inflammation is associated with a disruption of anabolic signals initiating muscle growth, it has been theorized that a signal-to-noise model may best describe the relationship between inflammation and muscle growth. By keeping the “noise” of chronic inflammation to a minimum, the localized acute inflammatory response signals a stronger anabolic response than could be achieved with higher levels of chronic inflammation.

MALIGNANCY

Malignancy is the tendency of a medical condition to become progressively worse. Malignancy is most familiar as a characterization of cancer. A *malignant* tumor contrasts with a non-cancerous *benign* tumor in that a malignancy is not self-limited in its growth, is capable of invading into adjacent tissues, and may be capable of spreading to distant tissues. A benign tumor has none of those properties. Malignancy in cancers is characterized by anaplasia, invasiveness, and metastasis. Malignant tumors are also characterized by genome instability, so that cancers, as assessed by whole genome sequencing, frequently have between 10,000 and 100,000 mutations in their entire genomes. Cancers usually show tumour heterogeneity, containing multiple subclones. They also frequently have reduced expression of DNA repair enzymes due to epigenetic methylation of DNA repair genes or altered microRNAs that control DNA repair gene expression.

Uses of “malignant” in oncology:

- Malignancy, malignant neoplasm and malignant tumor are synonymous with cancer
- Malignant ascites
- Malignant transformation

Non-oncologic disorders referred to as “malignant”:

- Malignant hypertension
- Malignant hyperthermia
- Malignant otitis externa
- Malignant tertian malaria (malaria caused specifically by *Plasmodium falciparum*)
- Neuroleptic malignant syndrome

NECROSIS

Necrosis is a form of cell injury which results in the premature death of cells in living tissue by autolysis. Necrosis is caused by factors external to the cell or tissue, such as infection, toxins, or trauma which result in the unregulated digestion of cell components.

In contrast, apoptosis is a naturally occurring programmed and targeted cause of cellular death.

While apoptosis often provides beneficial effects to the organism, necrosis is almost always detrimental and can be fatal. Cellular death due to necrosis does not follow the apoptotic signal transduction pathway, but rather various receptors are activated, and result in the loss of cell membrane integrity and an uncontrolled release of products of cell death into the extracellular space. This initiates in the surrounding tissue an inflammatory response which attracts leukocytes and nearby phagocytes which eliminate the dead cells by phagocytosis. However, microbial damaging substances released by leukocytes would create collateral damage to surrounding tissues. This excess collateral damage inhibits the healing process. Thus, untreated necrosis results in a build-up of decomposing dead tissue and cell debris at or near the site of the cell death. A classic example is gangrene. For this reason, it is often necessary to remove necrotic tissue surgically, a procedure known as debridement.

Classification

Structural signs that indicate irreversible cell injury and the progression of necrosis include dense clumping and progressive disruption of genetic material, and disruption to membranes of cells and organelles.

Morphological Patterns

There are six distinctive morphological patterns of necrosis:

- Coagulative necrosis is characterized by the formation of a gelatinous (gel-like) substance in dead tissues in which the architecture of the tissue is maintained, and can be observed by light microscopy. Coagulation occurs as a result of protein denaturation, causing albumin to transform into a firm and opaque state. This pattern of necrosis is typically seen in hypoxic (low-oxygen) environments, such as infarction. Coagulative necrosis occurs primarily in tissues such as the kidney, heart and adrenal glands. Severe ischemia most commonly causes necrosis of this form.
- Liquefactive necrosis (or colliquative necrosis), in contrast to coagulative necrosis, is characterized by the digestion of dead cells to form a viscous liquid mass. This is typical of bacterial, or sometimes fungal, infections because of their ability to stimulate an inflammatory response. The necrotic liquid mass is frequently creamy yellow due to the presence of dead leukocytes and is commonly known as pus. Hypoxic infarcts in the brain presents as this type of necrosis, because the brain contains little connective tissue but high amounts of digestive enzymes and lipids, and cells therefore can be readily digested by their own enzymes.
- Gangrenous necrosis can be considered a type of coagulative necrosis that resembles mummified tissue. It is characteristic of ischemia of lower limb and the gastrointestinal tracts. If superimposed infection of dead tissues occurs, then liquefactive necrosis ensues (wet gangrene)
- Caseous necrosis can be considered a combination of coagulative and liquefactive necrosis, typically caused by mycobacteria (*e.g.* tuberculosis), fungi and some foreign substances. The necrotic tissue appears as white and friable, like clumped cheese. Dead cells disintegrate but are not completely digested, leaving granular particles. Microscopic examination shows amorphous granular debris enclosed within a distinctive inflammatory border. Granuloma has this characteristic.
- Fat necrosis is specialized necrosis of fat tissue, resulting from the action of activated lipases on fatty tissues such as the pancreas. In the pancreas it leads to acute pancreatitis, a condition where the

pancreatic enzymes leak out into the peritoneal cavity, and liquefy the membrane by splitting the triglyceride esters into fatty acids through fat saponification. Calcium, magnesium or sodium may bind to these lesions to produce a chalky-white substance. The calcium deposits are microscopically distinctive and may be large enough to be visible on radiographic examinations. To the naked eye, calcium deposits appear as gritty white flecks.

- Fibrinoid necrosis is a special form of necrosis usually caused by immune-mediated vascular damage. It is marked by complexes of antigen and antibodies, sometimes referred to as “immune complexes” deposited within arterial walls together with fibrin.

Other Clinical Classifications of Necrosis

- There are also very specific forms of necrosis such as gangrene (term used in clinical practices for limbs which have suffered severe hypoxia), gummatous necrosis (due to spirochaetal infections) and hemorrhagic necrosis (due to the blockage of venous drainage of an organ or tissue).
- Some spider bites may lead to necrosis. In the United States, only spider bites from the brown recluse spider (genus *Loxosceles*) reliably progress to necrosis. In other countries, spiders of the same genus, such as the Chilean recluse in South America, are also known to cause necrosis. Claims that yellow sac spiders and hobo spiders possess necrotic venom have not been substantiated.
- In blind mole rats (genus *Spalax*), the process of necrosis replaces the role of the systematic apoptosis normally used in many organisms. Low oxygen conditions, such as those common in blind mole rats' burrows, usually cause cells to undergo apoptosis. In adaptation to higher tendency of cell death, blind mole rats evolved a mutation in the tumor suppressor protein p53 (which is also used in humans) to prevent cells from undergoing apoptosis. Human cancer patients have similar mutations, and blind mole rats were thought to be more susceptible to cancer because their cells cannot undergo apoptosis. However, after a specific amount of time (within 3 days according to a study conducted at the University of Rochester), the cells in blind mole rats release interferon-beta (which the immune system normally uses to counter viruses) in response to over-proliferation of cells caused by the suppression of apoptosis. In this case, the interferon-beta triggers cells to undergo necrosis, and this mechanism also kills cancer cells in blind mole rats. Because of tumor suppression mechanisms such as this, blind mole rats and other spalacids are resistant to cancer.

CAUSES

Necrosis may occur due to external or internal factors.

External factors may involve mechanical trauma (physical damage to the body which causes cellular breakdown), damage to blood vessels (which may disrupt blood supply to associated tissue), and ischemia. Thermal effects (extremely high or low temperature) can result in necrosis due to the disruption of cells. In frostbite, crystals form, increasing the pressure of remaining tissue and fluid causing the cells to burst. Under extreme conditions tissues and cells die through an unregulated process of destruction of membranes and cytosol. Internal factors causing necrosis include: trophoneurotic disorders; injury and paralysis of nerve cells. Pancreatic enzymes (lipases) are the major cause of fat necrosis. Necrosis can be activated by components of the immune system, such as the complement system; bacterial toxins; activated natural killer cells; and peritoneal macrophages.

Pathogen-induced necrosis programmes in cells with immunological barriers (intestinal mucosa) may alleviate invasion of pathogens through surfaces affected by inflammation. Toxins and pathogens may cause necrosis; toxins such as snake venoms may inhibit enzymes and cause cell death. Necrotic wounds have also resulted from the stings of *Vespa mandarinia*.

Pathological conditions are characterized by inadequate secretion of cytokines. Nitric oxide (NO) and reactive oxygen species (ROS) are also accompanied by intense necrotic death of cells. A classic example of a necrotic condition is ischemia which leads to a drastic depletion of oxygen, glucose, and other trophic factors and induces massive necrotic death of endothelial cells and non-proliferating cells of surrounding tissues (neurons, cardiomyocytes, renal cells, etc.). Recent cytological data indicates that necrotic death occurs not only during pathological events but it is also a component of some physiological process. Activation-induced death of primary T-lymphocytes and other important constituents of the immune response are caspase-independent and necrotic by morphology; hence, current researchers have demonstrated that the occurrence of necrotic cell death can not only occur during pathological processes but also during normal processes such as tissue renewal, embryogenesis, and immune response.

PATHOGENESIS

Until recently, necrosis was thought to be an unregulated process. There are two broad pathways in which necrosis may occur in an organism. The first of these two pathways initially involves oncosis, where swelling of the cells occur. The cell then proceeds to blebbing, and this is followed by pyknosis, in which nuclear shrinkage transpires. In the final step of this pathway the nucleus is dissolved into the cytoplasm, which is referred to as karyolysis.

The second pathway is a secondary form of necrosis that is shown to occur after apoptosis and budding. Cellular changes of necrosis occur in this secondary form of apoptosis, where the nucleus breaks into fragments, which is known as karyorrhexis.

Cellular Changes

The nucleus changes in necrosis, and characteristics of this change are determined by manner in which its DNA breaks down:

- *Karyolysis*: the chromatin of the nucleus fades due to the loss of the DNA by degradation.
- *Pyknosis*: the nucleus shrinks and the chromatin condenses.
- *Karyorrhexis*: the shrunken nucleus fragments to complete dispersal.

Plasma alterations are also seen in necrosis. Plasma membranes appear discontinuous when viewed with an electron microscope. This discontinuous membrane is caused by cell blebbing and the loss of microvilli.

TREATMENT

There are many causes of necrosis, and as such treatment is based upon how the necrosis came about. Treatment of necrosis typically involves two distinct processes: Usually, the underlying cause of the necrosis must be treated before the dead tissue itself can be dealt with.

- Debridement, referring to the removal of dead tissue by surgical or non-surgical means, is the standard therapy for necrosis. Depending on the severity of the necrosis, this may range from removal of small patches of skin to complete amputation of affected limbs or organs. Chemical removal of necrotic tissue is another option in which enzymatic debriding agents, categorised as proteolytic, fibrinolytic or collagenases, are used to target the various components of dead tissue. In select cases, special maggot therapy using *Lucilia sericata* larvae has been employed to remove necrotic tissue and infection.
- In the case of ischemia, which includes myocardial infarction, the restriction of blood supply to tissues causes hypoxia and the creation of reactive oxygen species (ROS) that react with, and damage proteins and membranes. Antioxidant treatments can be applied to scavenge the ROS.

- Wounds caused by physical agents, including physical trauma and chemical burns, can be treated with antibiotics and anti-inflammatory drugs to prevent bacterial infection and inflammation. Keeping the wound clean from infection also prevents necrosis.
- Chemical and toxic agents (*e.g.* pharmaceutical drugs, acids, bases) react with the skin leading to skin loss and eventually necrosis. Treatment involves identification and discontinuation of the harmful agent, followed by treatment of the wound, including prevention of infection and possibly the use of immunosuppressive therapies such as anti-inflammatory drugs or immunosuppressants. In the example of a snake bite, the use of anti-venom halts the spread of toxins whilst receiving antibiotics to impede infection.

Even after the initial cause of the necrosis has been halted, the necrotic tissue will remain in the body. The body's immune response to apoptosis, which involves the automatic breaking down and recycling of cellular material, is not triggered by necrotic cell death due to the apoptotic pathway being disabled.

IN PLANTS

If calcium is deficient, pectin cannot be synthesized, and therefore the cell walls cannot be bonded and thus an impediment of the meristems. This will lead to necrosis of stem and root tips and leaf edges. For example, necrosis of tissue can occur in *Arabidopsis thaliana* due to plant pathogens.

IMMUNE SYSTEM

The immune system is a host defence system comprising many biological structures and processes within an organism that protects against disease. To function properly, an immune system must detect a wide variety of agents, known as pathogens, from viruses to parasitic worms, and distinguish them from the organism's own healthy tissue. In many species, the immune system can be classified into subsystems, such as the innate immune system versus the adaptive immune system, or humoral immunity versus cell-mediated immunity. In humans, the blood–brain barrier, blood–cerebrospinal fluid barrier, and similar fluid–brain barriers separate the peripheral immune system from the neuroimmune system, which protects the brain. Pathogens can rapidly evolve and adapt, and thereby avoid detection and neutralization by the immune system; however, multiple defence mechanisms have also evolved to recognize and neutralize pathogens. Even simple unicellular organisms such as bacteria possess a rudimentary immune system in the form of enzymes that protect against bacteriophage infections. Other basic immune mechanisms evolved in ancient eukaryotes and remain in their modern descendants, such as plants and invertebrates. These mechanisms include phagocytosis, antimicrobial peptides called defensins, and the complement system. Jawed vertebrates, including humans, have even more sophisticated defence mechanisms, including the ability to adapt over time to recognize specific pathogens more efficiently. Adaptive (or acquired) immunity creates immunological memory after an initial response to a specific pathogen, leading to an enhanced response to subsequent encounters with that same pathogen. This process of acquired immunity is the basis of vaccination.

Disorders of the immune system can result in autoimmune diseases, inflammatory diseases and cancer. Immunodeficiency occurs when the immune system is less active than normal, resulting in recurring and life-threatening infections. In humans, immunodeficiency can either be the result of a genetic disease such as severe combined immunodeficiency, acquired conditions such as HIV/AIDS, or the use of immunosuppressive medication. In contrast, autoimmunity results from a hyperactive immune system attacking normal tissues as if they were foreign organisms. Common autoimmune diseases include Hashimoto's thyroiditis, rheumatoid arthritis, diabetes mellitus type 1, and systemic lupus erythematosus. Immunology covers the study of all aspects of the immune system.

HISTORY OF IMMUNOLOGY

Immunology is a science that examines the structure and function of the immune system. It originates from medicine and early studies on the causes of immunity to disease. The earliest known reference to immunity was during the plague of Athens in 430 BC. Thucydides noted that people who had recovered from a previous bout of the disease could nurse the sick without contracting the illness a second time. In the 18th century, Pierre-Louis Moreau de Maupertuis made experiments with scorpion venom and observed that certain dogs and mice were immune to this venom. This and other observations of acquired immunity were later exploited by Louis Pasteur in his development of vaccination and his proposed germ theory of disease. Pasteur's theory was in direct opposition to contemporary theories of disease, such as the miasma theory. It was not until Robert Koch's 1891 proofs, for which he was awarded a Nobel Prize in 1905, that microorganisms were confirmed as the cause of infectious disease. Viruses were confirmed as human pathogens in 1901, with the discovery of the yellow fever virus by Walter Reed. Immunology made a great advance towards the end of the 19th century, through rapid developments, in the study of humoral immunity and cellular immunity. Particularly important was the work of Paul Ehrlich, who proposed the side-chain theory to explain the specificity of the antigen-antibody reaction; his contributions to the understanding of humoral immunity were recognized by the award of a Nobel Prize in 1908, which was jointly awarded to the founder of cellular immunology, Elie Metchnikoff.

LAYERED DEFENCE

The immune system protects organisms from infection with layered defenses of increasing specificity. In simple terms, physical barriers prevent pathogens such as bacteria and viruses from entering the organism. If a pathogen breaches these barriers, the innate immune system provides an immediate, but non-specific response. Innate immune systems are found in all plants and animals. If pathogens successfully evade the innate response, vertebrates possess a second layer of protection, the adaptive immune system, which is activated by the innate response. Here, the immune system adapts its response during an infection to improve its recognition of the pathogen. This improved response is then retained after the pathogen has been eliminated, in the form of an immunological memory, and allows the adaptive immune system to mount faster and stronger attacks each time this pathogen is encountered. Both innate and adaptive immunity depend on the ability of the immune system to distinguish between self and non-self molecules. In immunology, *self* molecules are those components of an organism's body that can be distinguished from foreign substances by the immune system. Conversely, *non-self* molecules are those recognized as foreign molecules. One class of non-self molecules are called antigens (short for *antibody generators*) and are defined as substances that bind to specific immune receptors and elicit an immune response.

Table. Components of the immune system.

Innate immune system	Adaptive immune system
Response is non-specific	Pathogen and antigen specific response
Exposure leads to immediate maximal response	Lag time between exposure and maximal response
Cell-mediated and humoral components	Cell-mediated and humoral components
No immunological memory	Exposure leads to immunological memory
Found in nearly all forms of life	Found only in jawed vertebrates

INNATE IMMUNE SYSTEM

Microorganisms or toxins that successfully enter an organism encounter the cells and mechanisms of the innate immune system. The innate response is usually triggered when microbes are identified by pattern recognition receptors, which recognize components that are conserved among broad groups of microorganisms,

or when damaged, injured or stressed cells send out alarm signals, many of which (but not all) are recognized by the same receptors as those that recognize pathogens. Innate immune defenses are non-specific, meaning these systems respond to pathogens in a generic way. This system does not confer long-lasting immunity against a pathogen. The innate immune system is the dominant system of host defence in most organisms.

Surface Barriers

Several barriers protect organisms from infection, including mechanical, chemical, and biological barriers. The waxy cuticle of most leaves, the exoskeleton of insects, the shells and membranes of externally deposited eggs, and skin are examples of mechanical barriers that are the first line of defence against infection. However, as organisms cannot be completely sealed from their environments, other systems act to protect body openings such as the lungs, intestines, and the genitourinary tract. In the lungs, coughing and sneezing mechanically eject pathogens and other irritants from the respiratory tract. The flushing action of tears and urine also mechanically expels pathogens, while mucus secreted by the respiratory and gastrointestinal tract serves to trap and entangle microorganisms. Chemical barriers also protect against infection. The skin and respiratory tract secrete antimicrobial peptides such as the β -defensins. Enzymes such as lysozyme and phospholipase A2 in saliva, tears, and breast milk are also antibacterials. Vaginal secretions serve as a chemical barrier following menarche, when they become slightly acidic, while semen contains defensins and zinc to kill pathogens. In the stomach, gastric acid and proteases serve as powerful chemical defenses against ingested pathogens. Within the genitourinary and gastrointestinal tracts, commensal flora serve as biological barriers by competing with pathogenic bacteria for food and space and, in some cases, by changing the conditions in their environment, such as pH or available iron. As a result of the symbiotic relationship between commensals and the immune system, the probability that pathogens will reach sufficient numbers to cause illness is reduced. However, since most antibiotics non-specifically target bacteria and do not affect fungi, oral antibiotics can lead to an “overgrowth” of fungi and cause conditions such as a vaginal candidiasis (a yeast infection). There is good evidence that re-introduction of probiotic flora, such as pure cultures of the lactobacilli normally found in unpasteurized yogurt, helps restore a healthy balance of microbial populations in intestinal infections in children and encouraging preliminary data in studies on bacterial gastroenteritis, inflammatory bowel diseases, urinary tract infection and post-surgical infections.

Inflammation

Inflammation is one of the first responses of the immune system to infection. The symptoms of inflammation are redness, swelling, heat, and pain, which are caused by increased blood flow into tissue. Inflammation is produced by eicosanoids and cytokines, which are released by injured or infected cells. Eicosanoids include prostaglandins that produce fever and the dilation of blood vessels associated with inflammation, and leukotrienes that attract certain white blood cells (leukocytes). Common cytokines include interleukins that are responsible for communication between white blood cells; chemokines that promote chemotaxis; and interferons that have anti-viral effects, such as shutting down protein synthesis in the host cell. Growth factors and cytotoxic factors may also be released. These cytokines and other chemicals recruit immune cells to the site of infection and promote healing of any damaged tissue following the removal of pathogens.

Complement System

The complement system is a biochemical cascade that attacks the surfaces of foreign cells. It contains over 20 different proteins and is named for its ability to “complement” the killing of pathogens by antibodies. Complement is the major humoral component of the innate immune response. Many species have complement

systems, including non-mammals like plants, fish, and some invertebrates. In humans, this response is activated by complement binding to antibodies that have attached to these microbes or the binding of complement proteins to carbohydrates on the surfaces of microbes. This recognition signal triggers a rapid killing response. The speed of the response is a result of signal amplification that occurs after sequential proteolytic activation of complement molecules, which are also proteases. After complement proteins initially bind to the microbe, they activate their protease activity, which in turn activates other complement proteases, and so on. This produces a catalytic cascade that amplifies the initial signal by controlled positive feedback. The cascade results in the production of peptides that attract immune cells, increase vascular permeability, and opsonize (coat) the surface of a pathogen, marking it for destruction. This deposition of complement can also kill cells directly by disrupting their plasma membrane.

Cellular Barriers

Leukocytes (white blood cells) act like independent, single-celled organisms and are the second arm of the innate immune system. The innate leukocytes include the phagocytes (macrophages, neutrophils, and dendritic cells), innate lymphoid cells, mast cells, eosinophils, basophils, and natural killer cells. These cells identify and eliminate pathogens, either by attacking larger pathogens through contact or by engulfing and then killing microorganisms. Innate cells are also important mediators in lymphoid organ development and the activation of the adaptive immune system.

Phagocytosis is an important feature of cellular innate immunity performed by cells called phagocytes that engulf, or eat, pathogens or particles. Phagocytes generally patrol the body searching for pathogens, but can be called to specific locations by cytokines. Once a pathogen has been engulfed by a phagocyte, it becomes trapped in an intracellular vesicle called a phagosome, which subsequently fuses with another vesicle called a lysosome to form a phagolysosome. The pathogen is killed by the activity of digestive enzymes or following a respiratory burst that releases free radicals into the phagolysosome. Phagocytosis evolved as a means of acquiring nutrients, but this role was extended in phagocytes to include engulfment of pathogens as a defence mechanism. Phagocytosis probably represents the oldest form of host defence, as phagocytes have been identified in both vertebrate and invertebrate animals. Neutrophils and macrophages are phagocytes that travel throughout the body in pursuit of invading pathogens. Neutrophils are normally found in the bloodstream and are the most abundant type of phagocyte, normally representing 50 per cent to 60 per cent of the total circulating leukocytes, and consisting of neutrophil-killer and neutrophil-cager subpopulations. During the acute phase of inflammation, particularly as a result of bacterial infection, neutrophils migrate towards the site of inflammation in a process called chemotaxis, and are usually the first cells to arrive at the scene of infection. Macrophages are versatile cells that reside within tissues and produce a wide array of chemicals including enzymes, complement proteins, and cytokines, while they can also act as scavengers that rid the body of worn-out cells and other debris, and as antigen-presenting cells that activate the adaptive immune system. Dendritic cells (DC) are phagocytes in tissues that are in contact with the external environment; therefore, they are located mainly in the skin, nose, lungs, stomach, and intestines.

They are named for their resemblance to neuronal dendrites, as both have many spine-like projections, but dendritic cells are in no way connected to the nervous system. Dendritic cells serve as a link between the bodily tissues and the innate and adaptive immune systems, as they present antigens to T cells, one of the key cell types of the adaptive immune system. Mast cells reside in connective tissues and mucous membranes, and regulate the inflammatory response. They are most often associated with allergy and anaphylaxis. Basophils and eosinophils are related to neutrophils. They secrete chemical mediators that are involved in defending against parasites and play a role in allergic reactions, such as asthma. Natural killer (NK cells) cells are leukocytes that attack and destroy tumor cells, or cells that have been infected by viruses.

Natural Killer Cells

Natural killer cells, or NK cells, are lymphocytes and a component of the innate immune system which does not directly attack invading microbes. Rather, NK cells destroy compromised host cells, such as tumor cells or virus-infected cells, recognizing such cells by a condition known as “missing self.” This term describes cells with low levels of a cell-surface marker called MHC I (major histocompatibility complex) – a situation that can arise in viral infections of host cells. They were named “natural killer” because of the initial notion that they do not require activation in order to kill cells that are “missing self.” For many years it was unclear how NK cells recognize tumor cells and infected cells. It is now known that the MHC makeup on the surface of those cells is altered and the NK cells become activated through recognition of “missing self”. Normal body cells are not recognized and attacked by NK cells because they express intact self MHC antigens. Those MHC antigens are recognized by killer cell immunoglobulin receptors (KIR) which essentially put the brakes on NK cells.

ADAPTIVE IMMUNE SYSTEM

The adaptive immune system evolved in early vertebrates and allows for a stronger immune response as well as immunological memory, where each pathogen is “remembered” by a signature antigen. The adaptive immune response is antigen-specific and requires the recognition of specific “non-self” antigens during a process called antigen presentation. Antigen specificity allows for the generation of responses that are tailored to specific pathogens or pathogen-infected cells. The ability to mount these tailored responses is maintained in the body by “memory cells”. Should a pathogen infect the body more than once, these specific memory cells are used to quickly eliminate it.

Lymphocytes

The cells of the adaptive immune system are special types of leukocytes, called lymphocytes. B cells and T cells are the major types of lymphocytes and are derived from hematopoietic stem cells in the bone marrow. B cells are involved in the humoral immune response, whereas T cells are involved in cell-mediated immune response. Both B cells and T cells carry receptor molecules that recognize specific targets. T cells recognize a “non-self” target, such as a pathogen, only after antigens (small fragments of the pathogen) have been processed and presented in combination with a “self” receptor called a major histocompatibility complex (MHC) molecule. There are two major subtypes of T cells: the killer T cell and the helper T cell. In addition there are regulatory T cells which have a role in modulating immune response. Killer T cells only recognize antigens coupled to Class I MHC molecules, while helper T cells and regulatory T cells only recognize antigens coupled to Class II MHC molecules. These two mechanisms of antigen presentation reflect the different roles of the two types of T cell. A third, minor subtype are the $\gamma\delta$ T cells that recognize intact antigens that are not bound to MHC receptors. The double-positive T cells are exposed to a wide variety of self-antigens in the thymus, in which iodine is necessary for its thymus development and activity. In contrast, the B cell antigen-specific receptor is an antibody molecule on the B cell surface, and recognizes whole pathogens without any need for antigen processing. Each lineage of B cell expresses a different antibody, so the complete set of B cell antigen receptors represent all the antibodies that the body can manufacture.

Killer T cells

Killer T cells are a sub-group of T cells that kill cells that are infected with viruses (and other pathogens), or are otherwise damaged or dysfunctional. As with B cells, each type of T cell recognizes a different antigen. Killer T cells are activated when their T-cell receptor (TCR) binds to this specific antigen in a complex with the MHC Class I receptor of another cell. Recognition of this MHC:antigen complex is aided by a co-receptor on

the T cell, called CD8. The T cell then travels throughout the body in search of cells where the MHC I receptors bear this antigen. When an activated T cell contacts such cells, it releases cytotoxins, such as perforin, which form pores in the target cell's plasma membrane, allowing ions, water and toxins to enter. The entry of another toxin called granulysin (a protease) induces the target cell to undergo apoptosis. T cell killing of host cells is particularly important in preventing the replication of viruses. T cell activation is tightly controlled and generally requires a very strong MHC/antigen activation signal, or additional activation signals provided by "helper" T cells.

Helper T cells

Helper T cells regulate both the innate and adaptive immune responses and help determine which immune responses the body makes to a particular pathogen. These cells have no cytotoxic activity and do not kill infected cells or clear pathogens directly. They instead control the immune response by directing other cells to perform these tasks. Helper T cells express T cell receptors (TCR) that recognize antigen bound to Class II MHC molecules. The MHC:antigen complex is also recognized by the helper cell's CD4 co-receptor, which recruits molecules inside the T cell (*e.g.*, Lck) that are responsible for the T cell's activation. Helper T cells have a weaker association with the MHC:antigen complex than observed for killer T cells, meaning many receptors (around 200–300) on the helper T cell must be bound by an MHC:antigen in order to activate the helper cell, while killer T cells can be activated by engagement of a single MHC:antigen molecule. Helper T cell activation also requires longer duration of engagement with an antigen-presenting cell. The activation of a resting helper T cell causes it to release cytokines that influence the activity of many cell types. Cytokine signals produced by helper T cells enhance the microbicidal function of macrophages and the activity of killer T cells. In addition, helper T cell activation causes an upregulation of molecules expressed on the T cell's surface, such as CD40 ligand (also called CD154), which provide extra stimulatory signals typically required to activate antibody-producing B cells.

Gamma delta T cells

Gamma delta T cells ($\gamma\delta$ T cells) possess an alternative T-cell receptor (TCR) as opposed to CD4+ and CD8+ ($\alpha\beta$) T cells and share the characteristics of helper T cells, cytotoxic T cells and NK cells. The conditions that produce responses from $\gamma\delta$ T cells are not fully understood. Like other 'unconventional' T cell subsets bearing invariant TCRs, such as CD1d-restricted Natural Killer T cells, $\gamma\delta$ T cells straddle the border between innate and adaptive immunity. On one hand, $\gamma\delta$ T cells are a component of adaptive immunity as they rearrange TCR genes to produce receptor diversity and can also develop a memory phenotype. On the other hand, the various subsets are also part of the innate immune system, as restricted TCR or NK receptors may be used as pattern recognition receptors. For example, large numbers of human V γ 9/V δ 2 T cells respond within hours to common molecules produced by microbes, and highly restricted V δ 1+ T cells in epithelia respond to stressed epithelial cells.

B Lymphocytes and Antibodies

A B cell identifies pathogens when antibodies on its surface bind to a specific foreign antigen. This antigen/antibody complex is taken up by the B cell and processed by proteolysis into peptides. The B cell then displays these antigenic peptides on its surface MHC class II molecules. This combination of MHC and antigen attracts a matching helper T cell, which releases lymphokines and activates the B cell. As the activated B cell then begins to divide, its offspring (plasma cells) secrete millions of copies of the antibody that recognizes this antigen. These antibodies circulate in blood plasma and lymph, bind to pathogens expressing the antigen and

mark them for destruction by complement activation or for uptake and destruction by phagocytes. Antibodies can also neutralize challenges directly, by binding to bacterial toxins or by interfering with the receptors that viruses and bacteria use to infect cells.

Alternative Adaptive Immune System

Evolution of the adaptive immune system occurred in an ancestor of the jawed vertebrates. Many of the classical molecules of the adaptive immune system (*e.g.*, immunoglobulins and T-cell receptors) exist only in jawed vertebrates. However, a distinct lymphocyte-derived molecule has been discovered in primitive jawless vertebrates, such as the lamprey and hagfish. These animals possess a large array of molecules called Variable lymphocyte receptors (VLRs) that, like the antigen receptors of jawed vertebrates, are produced from only a small number (one or two) of genes. These molecules are believed to bind pathogenic antigens in a similar way to antibodies, and with the same degree of specificity.

Immunological Memory

When B cells and T cells are activated and begin to replicate, some of their offspring become long-lived memory cells. Throughout the lifetime of an animal, these memory cells remember each specific pathogen encountered and can mount a strong response if the pathogen is detected again. This is “adaptive” because it occurs during the lifetime of an individual as an adaptation to infection with that pathogen and prepares the immune system for future challenges. Immunological memory can be in the form of either passive short-term memory or active long-term memory.

Passive Memory

Newborn infants have no prior exposure to microbes and are particularly vulnerable to infection. Several layers of passive protection are provided by the mother. During pregnancy, a particular type of antibody, called IgG, is transported from mother to baby directly through the placenta, so human babies have high levels of antibodies even at birth, with the same range of antigen specificities as their mother. Breast milk or colostrum also contains antibodies that are transferred to the gut of the infant and protect against bacterial infections until the newborn can synthesize its own antibodies. This is passive immunity because the fetus does not actually make any memory cells or antibodies—it only borrows them. This passive immunity is usually short-term, lasting from a few days up to several months. In medicine, protective passive immunity can also be transferred artificially from one individual to another via antibody-rich serum.

Active Memory and Immunization

Long-term *active* memory is acquired following infection by activation of B and T cells. Active immunity can also be generated artificially, through vaccination. The principle behind vaccination (also called immunization) is to introduce an antigen from a pathogen in order to stimulate the immune system and develop specific immunity against that particular pathogen without causing disease associated with that organism. This deliberate induction of an immune response is successful because it exploits the natural specificity of the immune system, as well as its inducibility. With infectious disease remaining one of the leading causes of death in the human population, vaccination represents the most effective manipulation of the immune system mankind has developed.

Most viral vaccines are based on live attenuated viruses, while many bacterial vaccines are based on acellular components of micro-organisms, including harmless toxin components. Since many antigens derived from

acellular vaccines do not strongly induce the adaptive response, most bacterial vaccines are provided with additional adjuvants that activate the antigen-presenting cells of the innate immune system and maximize immunogenicity.

DISORDERS OF HUMAN IMMUNITY

The immune system is a remarkably effective structure that incorporates specificity, inducibility and adaptation. Failures of host defence do occur, however, and fall into three broad categories: immunodeficiencies, autoimmunity, and hypersensitivities.

Immunodeficiencies

Immunodeficiencies occur when one or more of the components of the immune system are inactive. The ability of the immune system to respond to pathogens is diminished in both the young and the elderly, with immune responses beginning to decline at around 50 years of age due to immunosenescence. In developed countries, obesity, alcoholism, and drug use are common causes of poor immune function. However, malnutrition is the most common cause of immunodeficiency in developing countries. Diets lacking sufficient protein are associated with impaired cell-mediated immunity, complement activity, phagocyte function, IgA antibody concentrations, and cytokine production. Additionally, the loss of the thymus at an early age through genetic mutation or surgical removal results in severe immunodeficiency and a high susceptibility to infection. Immunodeficiencies can also be inherited or ‘acquired’. Chronic granulomatous disease, where phagocytes have a reduced ability to destroy pathogens, is an example of an inherited, or congenital, immunodeficiency. AIDS and some types of cancer cause acquired immunodeficiency.

Autoimmunity

Overactive immune responses comprise the other end of immune dysfunction, particularly the autoimmune disorders. Here, the immune system fails to properly distinguish between self and non-self, and attacks part of the body. Under normal circumstances, many T cells and antibodies react with “self” peptides. One of the functions of specialized cells (located in the thymus and bone marrow) is to present young lymphocytes with self antigens produced throughout the body and to eliminate those cells that recognize self-antigens, preventing autoimmunity.

Hypersensitivity

Hypersensitivity is an immune response that damages the body’s own tissues. They are divided into four classes (Type I – IV) based on the mechanisms involved and the time course of the hypersensitive reaction. Type I hypersensitivity is an immediate or anaphylactic reaction, often associated with allergy. Symptoms can range from mild discomfort to death. Type I hypersensitivity is mediated by IgE, which triggers degranulation of mast cells and basophils when cross-linked by antigen. Type II hypersensitivity occurs when antibodies bind to antigens on the patient’s own cells, marking them for destruction. This is also called antibody-dependent (or cytotoxic) hypersensitivity, and is mediated by IgG and IgM antibodies. Immune complexes (aggregations of antigens, complement proteins, and IgG and IgM antibodies) deposited in various tissues trigger Type III hypersensitivity reactions. Type IV hypersensitivity (also known as cell-mediated or *delayed type hypersensitivity*) usually takes between two and three days to develop. Type IV reactions are involved in many autoimmune and infectious diseases, but may also involve *contact dermatitis* (poison ivy). These reactions are mediated by T cells, monocytes, and macrophages.

Idiopathic Inflammation

Inflammation is one of the first responses of the immune system to infection, but it can appear without known cause. Inflammation is produced by eicosanoids and cytokines, which are released by injured or infected cells. Eicosanoids include prostaglandins that produce fever and the dilation of blood vessels associated with inflammation, and leukotrienes that attract certain white blood cells (leukocytes). Common cytokines include interleukins that are responsible for communication between white blood cells; chemokines that promote chemotaxis; and interferons that have anti-viral effects, such as shutting down protein synthesis in the host cell. Growth factors and cytotoxic factors may also be released. These cytokines and other chemicals recruit immune cells to the site of infection and promote healing of any damaged tissue following the removal of pathogens.

OTHER MECHANISMS AND EVOLUTION

It is likely that a multicomponent, adaptive immune system arose with the first vertebrates, as invertebrates do not generate lymphocytes or an antibody-based humoral response. Many species, however, utilize mechanisms that appear to be precursors of these aspects of vertebrate immunity. Immune systems appear even in the structurally most simple forms of life, with bacteria using a unique defence mechanism, called the restriction modification system to protect themselves from viral pathogens, called bacteriophages. Prokaryotes also possess acquired immunity, through a system that uses CRISPR sequences to retain fragments of the genomes of phage that they have come into contact with in the past, which allows them to block virus replication through a form of RNA interference. Prokaryotes also possess other defence mechanisms. Offensive elements of the immune systems are also present in unicellular eukaryotes, but studies of their roles in defence are few. Pattern recognition receptors are proteins used by nearly all organisms to identify molecules associated with pathogens. Antimicrobial peptides called defensins are an evolutionarily conserved component of the innate immune response found in all animals and plants, and represent the main form of invertebrate systemic immunity. The complement system and phagocytic cells are also used by most forms of invertebrate life. Ribonucleases and the RNA interference pathway are conserved across all eukaryotes, and are thought to play a role in the immune response to viruses. Unlike animals, plants lack phagocytic cells, but many plant immune responses involve systemic chemical signals that are sent through a plant. Individual plant cells respond to molecules associated with pathogens known as Pathogen-associated molecular patterns or PAMPs. When a part of a plant becomes infected, the plant produces a localized hypersensitive response, whereby cells at the site of infection undergo rapid apoptosis to prevent the spread of the disease to other parts of the plant. Systemic acquired resistance (SAR) is a type of defensive response used by plants that renders the entire plant resistant to a particular infectious agent. RNA silencing mechanisms are particularly important in this systemic response as they can block virus replication.

TUMOR IMMUNOLOGY

Another important role of the immune system is to identify and eliminate tumors. This is called immune surveillance. The *transformed cells* of tumors express antigens that are not found on normal cells. To the immune system, these antigens appear foreign, and their presence causes immune cells to attack the transformed tumor cells.

The antigens expressed by tumors have several sources; some are derived from oncogenic viruses like human papillomavirus, which causes cervical cancer, while others are the organism's own proteins that occur at low levels in normal cells but reach high levels in tumor cells. One example is an enzyme called tyrosinase that, when expressed at high levels, transforms certain skin cells (*e.g.* melanocytes) into tumors called melanomas. A third possible source of tumor antigens are proteins normally important for regulating cell growth and survival, that commonly mutate into cancer inducing molecules called oncogenes.

The main response of the immune system to tumors is to destroy the abnormal cells using killer T cells, sometimes with the assistance of helper T cells. Tumor antigens are presented on MHC class I molecules in a similar way to viral antigens. This allows killer T cells to recognize the tumor cell as abnormal. NK cells also kill tumorous cells in a similar way, especially if the tumor cells have fewer MHC class I molecules on their surface than normal; this is a common phenomenon with tumors. Sometimes antibodies are generated against tumor cells allowing for their destruction by the complement system. Clearly, some tumors evade the immune system and go on to become cancers. Tumor cells often have a reduced number of MHC class I molecules on their surface, thus avoiding detection by killer T cells. Some tumor cells also release products that inhibit the immune response; for example by secreting the cytokine TGF- β , which suppresses the activity of macrophages and lymphocytes. In addition, immunological tolerance may develop against tumor antigens, so the immune system no longer attacks the tumor cells. Paradoxically, macrophages can promote tumor growth when tumor cells send out cytokines that attract macrophages, which then generate cytokines and growth factors such as tumor-necrosis factor alpha that nurture tumor development or promote stem-cell-like plasticity. In addition, a combination of hypoxia in the tumor and a cytokine produced by macrophages induces tumor cells to decrease production of a protein that blocks metastasis and thereby assists spread of cancer cells.

PHYSIOLOGICAL REGULATION

The immune system is involved in many aspects of physiological regulation in the body. The immune system interacts intimately with other systems, such as the endocrine and the nervous systems. The immune system also plays a crucial role in embryogenesis (development of the embryo), as well as in tissue repair and regeneration.

Hormones

Hormones can act as immunomodulators, altering the sensitivity of the immune system. For example, female sex hormones are known immunostimulators of both adaptive and innate immune responses. Some autoimmune diseases such as lupus erythematosus strike women preferentially, and their onset often coincides with puberty. By contrast, male sex hormones such as testosterone seem to be immunosuppressive. Other hormones appear to regulate the immune system as well, most notably prolactin, growth hormone and vitamin D.

Vitamin D

When a T-cell encounters a foreign pathogen, it extends a vitamin D receptor. This is essentially a signaling device that allows the T-cell to bind to the active form of vitamin D, the steroid hormone calcitriol. T-cells have a symbiotic relationship with vitamin D. Not only does the T-cell extend a vitamin D receptor, in essence asking to bind to the steroid hormone version of vitamin D, calcitriol, but the T-cell expresses the gene CYP27B1, which is the gene responsible for converting the pre-hormone version of vitamin D, calcidiol into the steroid hormone version, calcitriol. Only after binding to calcitriol can T-cells perform their intended function. Other immune system cells that are known to express CYP27B1 and thus activate vitamin D calcidiol, are dendritic cells, keratinocytes and macrophages.

It is conjectured that a progressive decline in hormone levels with age is partially responsible for weakened immune responses in aging individuals. Conversely, some hormones are regulated by the immune system, notably thyroid hormone activity. The age-related decline in immune function is also related to decreasing vitamin D levels in the elderly. As people age, two things happen that negatively affect their vitamin D levels. First, they stay indoors more due to decreased activity levels. This means that they get less sun and therefore produce less cholecalciferol via UVB radiation. Second, as a person ages the skin becomes less adept at producing vitamin D.

Sleep and Rest

The immune system is affected by sleep and rest, and sleep deprivation is detrimental to immune function. Complex feedback loops involving cytokines, such as interleukin-1 and tumor necrosis factor- α produced in response to infection, appear to also play a role in the regulation of non-rapid eye movement (REM) sleep. Thus the immune response to infection may result in changes to the sleep cycle, including an increase in slow-wave sleep relative to REM sleep. When suffering from sleep deprivation, active immunizations may have a diminished effect and may result in lower antibody production, and a lower immune response, than would be noted in a well-rested individual. Additionally, proteins such as NFIL3, which have been shown to be closely intertwined with both T-cell differentiation and our circadian rhythms, can be affected through the disturbance of natural light and dark cycles through instances of sleep deprivation, shift work, etc. As a result, these disruptions can lead to an increase in chronic conditions such as heart disease, chronic pain, and asthma. In addition to the negative consequences of sleep deprivation, sleep and the intertwined circadian system have been shown to have strong regulatory effects on immunological functions affecting both the innate and the adaptive immunity. First, during the early slow-wave-sleep stage, a sudden drop in blood levels of cortisol, epinephrine, and norepinephrine induce increased blood levels of the hormones leptin, pituitary growth hormone, and prolactin. These signals induce a pro-inflammatory state through the production of the pro-inflammatory cytokines interleukin-1, interleukin-12, TNF- α and IFN- γ . These cytokines then stimulate immune functions such as immune cells activation, proliferation, and differentiation. It is during this time that undifferentiated, or less differentiated, like naïve and central memory T cells, peak (*i.e.* during a time of a slowly evolving adaptive immune response). In addition to these effects, the milieu of hormones produced at this time (leptin, pituitary growth hormone, and prolactin) support the interactions between APCs and T-cells, a shift of the T_h1/T_h2 cytokine balance towards one that supports T_h1 , an increase in overall T_h cell proliferation, and naïve T cell migration to lymph nodes.

This milieu is also thought to support the formation of long-lasting immune memory through the initiation of $Th1$ immune responses. In contrast, during wake periods differentiated effector cells, such as cytotoxic natural killer cells and CTLs (cytotoxic T lymphocytes), peak in order to elicit an effective response against any intruding pathogens. As well during awake active times, anti-inflammatory molecules, such as cortisol and catecholamines, peak. There are two theories as to why the pro-inflammatory state is reserved for sleep time. First, inflammation would cause serious cognitive and physical impairments if it were to occur during wake times. Second, inflammation may occur during sleep times due to the presence of melatonin. Inflammation causes a great deal of oxidative stress and the presence of melatonin during sleep times could actively counteract free radical production during this time.

Nutrition and Diet

Overnutrition is associated with diseases such as diabetes and obesity, which are known to affect immune function. More moderate malnutrition, as well as certain specific trace mineral and nutrient deficiencies, can also compromise the immune response. Foods rich in certain fatty acids may foster a healthy immune system. Likewise, fetal undernourishment can cause a lifelong impairment of the immune system.

Repair and Regeneration

The immune system, particularly the innate component, plays a decisive role in tissue repair after an insult. Key actors include macrophages and neutrophils, but other cellular actors, including $\gamma\delta$ T cells, innate lymphoid cells (ILCs), and regulatory T cells (Tregs), are also important. The plasticity of immune cells and the balance between pro-inflammatory and anti-inflammatory signals are crucial aspects of efficient tissue repair. Immune

components and pathways are involved in regeneration as well, for example in amphibians. According to one hypothesis, organisms that can regenerate could be less immunocompetent than organisms that cannot regenerate.

MANIPULATION IN MEDICINE

The immune response can be manipulated to suppress unwanted responses resulting from autoimmunity, allergy, and transplant rejection, and to stimulate protective responses against pathogens that largely elude the immune system or cancer.

Immunosuppression

Immunosuppressive drugs are used to control autoimmune disorders or inflammation when excessive tissue damage occurs, and to prevent transplant rejection after an organ transplant. Anti-inflammatory drugs are often used to control the effects of inflammation. Glucocorticoids are the most powerful of these drugs; however, these drugs can have many undesirable side effects, such as central obesity, hyperglycemia, osteoporosis, and their use must be tightly controlled. Lower doses of anti-inflammatory drugs are often used in conjunction with cytotoxic or immunosuppressive drugs such as methotrexate or azathioprine. Cytotoxic drugs inhibit the immune response by killing dividing cells such as activated T cells. However, the killing is indiscriminate and other constantly dividing cells and their organs are affected, which causes toxic side effects. Immunosuppressive drugs such as cyclosporin prevent T cells from responding to signals correctly by inhibiting signal transduction pathways.

Immunostimulation

Cancer immunotherapy covers the medical ways to stimulate the immune system to attack cancer tumours.

THEORETICAL APPROACHES TO THE IMMUNE SYSTEM

Immunology is strongly experimental in everyday practice but is also characterized by an ongoing theoretical attitude. Many theories have been suggested in immunology from the end of the nineteenth century up to the present time. The end of the 19th century and the beginning of the 20th century saw a battle between “cellular” and “humoral” theories of immunity. According to the cellular theory of immunity, represented in particular by Elie Metchnikoff, it was cells – more precisely, phagocytes – that were responsible for immune responses. In contrast, the humoral theory of immunity, held, among others, by Robert Koch and Emil von Behring, stated that the active immune agents were soluble components (molecules) found in the organism’s “humors” rather than its cells. In the mid-1950s, Frank Burnet, inspired by a suggestion made by Niels Jerne, formulated the clonal selection theory (CST) of immunity. On the basis of CST, Burnet developed a theory of how an immune response is triggered according to the self/nonself distinction: “self” constituents (constituents of the body) do not trigger destructive immune responses, while “nonself” entities (pathogens, an allograft) trigger a destructive immune response. The theory was later modified to reflect new discoveries regarding histocompatibility or the complex “two-signal” activation of T cells. The self/nonself theory of immunity and the self/nonself vocabulary have been criticized, but remain very influential. More recently, several theoretical frameworks have been suggested in immunology, including “autopoietic” views, “cognitive immune” views, the “danger model” (or “danger theory”), and the “discontinuity” theory. The danger model, suggested by Polly Matzinger and colleagues, has been very influential, arousing many comments and discussions.

PREDICTING IMMUNOGENICITY

Larger drugs (>500 Da) can provoke a neutralizing immune response, particularly if the drugs are administered repeatedly, or in larger doses. This limits the effectiveness of drugs based on larger peptides and proteins

(which are typically larger than 6000 Da). In some cases, the drug itself is not immunogenic, but may be co-administered with an immunogenic compound, as is sometimes the case for Taxol. Computational methods have been developed to predict the immunogenicity of peptides and proteins, which are particularly useful in designing therapeutic antibodies, assessing likely virulence of mutations in viral coat particles, and validation of proposed peptide-based drug treatments. Early techniques relied mainly on the observation that hydrophilic amino acids are overrepresented in epitope regions than hydrophobic amino acids; however, more recent developments rely on machine learning techniques using databases of existing known epitopes, usually on well-studied virus proteins, as a training set. A publicly accessible database has been established for the cataloguing of epitopes from pathogens known to be recognizable by B cells. The emerging field of bioinformatics-based studies of immunogenicity is referred to as *immunoinformatics*. Immunoproteomics is the study of large sets of proteins (proteomics) involved in the immune response.

MANIPULATION BY PATHOGENS

The success of any pathogen depends on its ability to elude host immune responses. Therefore, pathogens evolved several methods that allow them to successfully infect a host, while evading detection or destruction by the immune system. Bacteria often overcome physical barriers by secreting enzymes that digest the barrier, for example, by using a type II secretion system. Alternatively, using a type III secretion system, they may insert a hollow tube into the host cell, providing a direct route for proteins to move from the pathogen to the host. These proteins are often used to shut down host defenses.

An evasion strategy used by several pathogens to avoid the innate immune system is to hide within the cells of their host (also called intracellular pathogenesis). Here, a pathogen spends most of its life-cycle inside host cells, where it is shielded from direct contact with immune cells, antibodies and complement. Some examples of intracellular pathogens include viruses, the food poisoning bacterium *Salmonella* and the eukaryotic parasites that cause malaria (*Plasmodium falciparum*) and leishmaniasis (*Leishmania spp.*). Other bacteria, such as *Mycobacterium tuberculosis*, live inside a protective capsule that prevents lysis by complement. Many pathogens secrete compounds that diminish or misdirect the host's immune response. Some bacteria form biofilms to protect themselves from the cells and proteins of the immune system. Such biofilms are present in many successful infections, e.g., the chronic *Pseudomonas aeruginosa* and *Burkholderia cenocepacia* infections characteristic of cystic fibrosis. Other bacteria generate surface proteins that bind to antibodies, rendering them ineffective; examples include *Streptococcus* (protein G), *Staphylococcus aureus* (protein A), and *Peptostreptococcus magnus* (protein L). The mechanisms used to evade the adaptive immune system are more complicated. The simplest approach is to rapidly change non-essential epitopes (amino acids and/or sugars) on the surface of the pathogen, while keeping essential epitopes concealed. This is called antigenic variation. An example is HIV, which mutates rapidly, so the proteins on its viral envelope that are essential for entry into its host target cell are constantly changing. These frequent changes in antigens may explain the failures of vaccines directed at this virus. The parasite *Trypanosoma brucei* uses a similar strategy, constantly switching one type of surface protein for another, allowing it to stay one step ahead of the antibody response. Masking antigens with host molecules is another common strategy for avoiding detection by the immune system. In HIV, the envelope that covers the virion is formed from the outermost membrane of the host cell; such “self-cloaked” viruses make it difficult for the immune system to identify them as “non-self” structures.

Adaptive Immune System

INTRODUCTION

The adaptive immune system, also known as the acquired immune system or, more rarely, as the specific immune system, is a subsystem of the overall immune system that is composed of highly specialized, systemic cells and processes that eliminate pathogens or prevent their growth. The adaptive immune system is one of the two main immunity strategies found in vertebrates (the other being the innate immune system). Adaptive immunity creates immunological memory after an initial response to a specific pathogen, and leads to an enhanced response to subsequent encounters with that pathogen. This process of acquired immunity is the basis of vaccination. Like the innate system, the adaptive system includes both humoral immunity components and cell-mediated immunity components. Unlike the innate immune system, the adaptive immune system is highly specific to a particular pathogen. Adaptive immunity can also provide long-lasting protection; for example, someone who recovers from measles is now protected against measles for their lifetime. In other cases it does not provide lifetime protection; for example, chickenpox. The adaptive system response destroys invading pathogens and any toxic molecules they produce. Sometimes the adaptive system is unable to distinguish harmful from harmless foreign molecules; the effects of this may be hayfever, asthma or any other allergy. Antigens are any substances that elicit the adaptive immune response. The cells that carry out the adaptive immune response are white blood cells known as lymphocytes. Two main broad classes—antibody responses and cell mediated immune response—are also carried by two different lymphocytes (B cells and T cells). In antibody responses, B cells are activated to secrete antibodies, which are proteins also known as immunoglobulins. Antibodies travel through the bloodstream and bind to the foreign antigen causing it to inactivate, which does not allow the antigen to bind to the host.

In acquired immunity, pathogen-specific receptors are “acquired” during the lifetime of the organism (whereas in innate immunity pathogen-specific receptors are already encoded in the germline). The acquired response is called “adaptive” because it prepares the body’s immune system for future challenges (though it can actually also be maladaptive when it results in autoimmunity). The system is highly adaptable because of somatic hypermutation (a process of accelerated somatic mutations), and V(D)J recombination (an irreversible genetic recombination of antigen receptor gene segments). This mechanism allows a small number of genes to generate a vast number of different antigen receptors, which are then uniquely expressed on each individual lymphocyte. Since the gene rearrangement leads to an irreversible change in the DNA of each cell, all progeny (offspring) of that cell inherit genes that encode the same receptor specificity, including the memory B cells and memory T cells that are the keys to long-lived specific immunity.

A theoretical framework explaining the workings of the adaptive immune system is provided by immune network theory. This theory, which builds on established concepts of clonal selection, is being applied in the search for an HIV vaccine.

FUNCTIONS

Adaptive immunity is triggered in vertebrates when a pathogen evades the innate immune system and (1) generates a threshold level of antigen and (2) generates “stranger” or “danger” signals activating dendritic cells.

The major functions of the adaptive immune system include:

- Recognition of specific “non-self” antigens in the presence of “self”, during the process of antigen presentation.
- Generation of responses that are tailored to maximally eliminate specific pathogens or pathogen-infected cells.
- Development of immunological memory, in which pathogens are “remembered” through memory B cells and memory T cells.

LYMPHOCYTES

A lymphocyte is one of the subtypes of white blood cell in a vertebrate’s immune system. Lymphocytes include natural killer cells (which function in cell-mediated, cytotoxic innate immunity), T cells (for cell-mediated, cytotoxic adaptive immunity), and B cells (for humoral, antibody-driven adaptive immunity). They are the main type of cell found in lymph, which prompted the name “lymphocyte”.

Types

The three major types of lymphocyte are T cells, B cells and natural killer (NK) cells. Lymphocytes can be identified by their large nucleus.

T cells and B Cells

T cells (thymus cells) and B cells (bone marrow- or bursa-derived cells) are the major cellular components of the adaptive immune response. T cells are involved in cell-mediated immunity, whereas B cells are primarily responsible for humoral immunity (relating to antibodies). The function of T cells and B cells is to recognize specific “non-self” antigens, during a process known as antigen presentation. Once they have identified an invader, the cells generate specific responses that are tailored to maximally eliminate specific pathogens or pathogen-infected cells. B cells respond to pathogens by producing large quantities of antibodies which then neutralize foreign objects like bacteria and viruses. In response to pathogens some T cells, called *T helper cells*, produce cytokines that direct the immune response, while other T cells, called *cytotoxic T cells*, produce toxic granules that contain powerful enzymes which induce the death of pathogen-infected cells. Following activation, B cells and T cells leave a lasting legacy of the antigens they have encountered, in the form of *memory cells*. Throughout the lifetime of an animal, these memory cells will remember each specific pathogen encountered, and are able to mount a strong and rapid response if the same pathogen is detected again; this is known as acquired immunity.

Natural Killer Cells

NK cells are a part of the innate immune system and play a major role in defending the host from both tumors and virally infected cells. NK cells distinguish infected cells and tumors from normal and uninfected cells by recognizing changes of a surface molecule called MHC (major histocompatibility complex) class I. NK cells are activated in response to a family of cytokines called interferons. Activated NK cells release cytotoxic (cell-killing) granules which then destroy the altered cells. They are named “natural killer cells” because they do not require prior activation in order to kill cells which are missing MHC class I.

DEVELOPMENT

haematopoiesis. All lymphocytes originate, during this process, from a common lymphoid progenitor before differentiating into their distinct lymphocyte types. The differentiation of lymphocytes follows various pathways in a hierarchical fashion as well as in a more plastic fashion. The formation of lymphocytes is known as lymphopoiesis. B cells mature into B lymphocytes in the bursa equivalent, which in humans is the GALT, which is thought to be located in the Peyer's patches of the intestine, while T cells migrate to and mature in a distinct organ, called the thymus. Following maturation, the lymphocytes enter the circulation and peripheral lymphoid organs (*e.g.* the spleen and lymph nodes) where they survey for invading pathogens and/or tumor cells. The lymphocytes involved in adaptive immunity (*i.e.* B and T cells) differentiate further after exposure to an antigen; they form effector and memory lymphocytes. Effector lymphocytes function to eliminate the antigen, either by releasing antibodies (in the case of B cells), cytotoxic granules (cytotoxic T cells) or by signaling to other cells of the immune system (helper T cells). Memory T cells remain in the peripheral tissues and circulation for an extended time ready to respond to the same antigen upon future exposure; they live weeks to several years, which is very long compared to other leukocytes.

CHARACTERISTICS

Microscopically, in a Wright's stained peripheral blood smear, a normal lymphocyte has a large, dark-staining nucleus with little to no eosinophilic cytoplasm. In normal situations, the coarse, dense nucleus of a lymphocyte is approximately the size of a red blood cell (about 7 μm in diameter). Some lymphocytes show a clear perinuclear zone (or halo) around the nucleus or could exhibit a small clear zone to one side of the nucleus. Polyribosomes are a prominent feature in the lymphocytes and can be viewed with an electron microscope. The ribosomes are involved in protein synthesis, allowing the generation of large quantities of cytokines and immunoglobulins by these cells. It is impossible to distinguish between T cells and B cells in a peripheral blood smear. Normally, flow cytometry testing is used for specific lymphocyte population counts. This can be used to determine the percentage of lymphocytes that contain a particular combination of specific cell surface proteins, such as immunoglobulins or cluster of differentiation (CD) markers or that produce particular proteins (for example, cytokines using intracellular cytokine staining (ICCS)). In order to study the function of a lymphocyte by virtue of the proteins it generates, other scientific techniques like the ELISPOT or secretion assay techniques can be used.

Table. Typical recognition markers for lymphocytes.

Class	Function	Proportion	Phenotypic marker(s)
Natural killer cells	Lysis of virally infected cells and tumour cells	7% (2–13%)	CD16 CD56 but not CD3
T helper cells	Release cytokines and growth factors that regulate other immune cells	46% (28–59%)	TCR $\alpha\beta$, CD3 and CD4
Cytotoxic T cells	Lysis of virally infected cells, tumour cells and allografts	19% (13–32%)	TCR $\alpha\beta$, CD3 and CD8
Gamma delta T cells	Immunoregulation and cytotoxicity	5% (2–8%)	TCR $\gamma\delta$ and CD3
B cells	Secretion of antibodies	23% (18–47%)	MHC class II, CD19 and CD20

In the circulatory system, they move from lymph node to lymph node. This contrasts with macrophages, which are rather stationary in the nodes.

LYMPHOCYTES AND DISEASE

A lymphocyte count is usually part of a peripheral complete blood cell count and is expressed as the percentage of lymphocytes to the total number of white blood cells counted. A general increase in the number of lymphocytes is known as lymphocytosis, whereas a decrease is known as lymphocytopenia.

High

An increase in lymphocyte concentration is usually a sign of a viral infection (in some rare case, leukemias are found through an abnormally raised lymphocyte count in an otherwise normal person). A high lymphocyte count with a low neutrophil count might be caused by lymphoma. Pertussis toxin (PTx) of *Bordetella pertussis*, formerly known as lymphocytosis-promoting factor, causes a decrease in the entry of lymphocytes into lymph nodes, which can lead to a condition known as lymphocytosis, with a complete lymphocyte count of over 4000 per μl in adults or over 8000 per μl in children. This is unique in that many bacterial infections illustrate neutrophil-predominance instead.

Low

A low normal to low absolute lymphocyte concentration is associated with increased rates of infection after surgery or trauma. One basis for low T cell lymphocytes occurs when the human immunodeficiency virus (HIV) infects and destroys T cells (specifically, the CD4 subgroup of T lymphocytes). Without the key defence that these T cells provide, the body becomes susceptible to opportunistic infections that otherwise would not affect healthy people. The extent of HIV progression is typically determined by measuring the percentage of CD4 T cells in the patient's blood – HIV ultimately progresses to acquired immune deficiency syndrome (AIDS). The effects of other viruses or lymphocyte disorders can also often be estimated by counting the numbers of lymphocytes present in the blood.

Tumor-infiltrating Lymphocytes

In some cancers, such as melanoma and colorectal cancer, lymphocytes can migrate into and attack the tumor. This can sometimes lead to regression of the primary tumor.

T LYMPHOCYTES

CD8+ T lymphocytes and cytotoxicity

Cytotoxic T cells (also known as TC, killer T cell, or cytotoxic T-lymphocyte (CTL)) are a sub-group of T cells that induce the death of cells that are infected with viruses (and other pathogens), or are otherwise damaged or dysfunctional. Naive cytotoxic T cells are activated when their T-cell receptor (TCR) strongly interacts with a peptide-bound MHC class I molecule. This affinity depends on the type and orientation of the antigen/MHC complex, and is what keeps the CTL and infected cell bound together. Once activated, the CTL undergoes a process called clonal selection, in which it gains functions and divides rapidly to produce an army of “armed” effector cells. Activated CTL then travels throughout the body searching for cells that bear that unique MHC Class I + peptide.

When exposed to these infected or dysfunctional somatic cells, effector CTL release perforin and granulysin: cytotoxins that form pores in the target cell's plasma membrane, allowing ions and water to flow into the infected cell, and causing it to burst or lyse. CTL release granzyme, a serine protease that enters cells via pores to induce apoptosis (cell death). To limit extensive tissue damage during an infection, CTL activation is tightly controlled and in general requires a very strong MHC/antigen activation signal, or additional activation signals provided by “helper” T-cells.

On resolution of the infection, most effector cells die and phagocytes clear them away—but a few of these cells remain as memory cells. On a later encounter with the same antigen, these memory cells quickly differentiate into effector cells, dramatically shortening the time required to mount an effective response.

HELPER T-CELLS

CD4⁺ lymphocytes, also called “helper” or “regulatory” T cells, are immune response mediators, and play an important role in establishing and maximizing the capabilities of the adaptive immune response. These cells have no cytotoxic or phagocytic activity; and cannot kill infected cells or clear pathogens, but, in essence “manage” the immune response, by directing other cells to perform these tasks. Helper T cells express T cell receptors (TCR) that recognize antigen bound to Class II MHC molecules. The activation of a naive helper T-cell causes it to release cytokines, which influences the activity of many cell types, including the APC (Antigen-Presenting Cell) that activated it. Helper T-cells require a much milder activation stimulus than cytotoxic T cells. Helper T cells can provide extra signals that “help” activate cytotoxic cells.

Th1 and Th2: Helper T Cell Responses

Classically, two types of effector CD4⁺ T helper cell responses can be induced by a professional APC, designated Th1 and Th2, each designed to eliminate different types of pathogens. The factors that dictate whether an infection triggers a Th1 or Th2 type response are not fully understood, but the response generated does play an important role in the clearance of different pathogens. The Th1 response is characterized by the production of Interferon-gamma, which activates the bactericidal activities of macrophages, and induces B cells to make opsonizing (coating) and complement-fixing antibodies, and leads to *cell-mediated immunity*. In general, Th1 responses are more effective against intracellular pathogens (viruses and bacteria that are inside host cells). The Th2 response is characterized by the release of Interleukin 5, which induces eosinophils in the clearance of parasites. Th2 also produce Interleukin 4, which facilitates B cell isotype switching. In general, Th2 responses are more effective against extracellular bacteria, parasites including helminths and toxins. Like cytotoxic T cells, most of the CD4⁺ helper cells die on resolution of infection, with a few remaining as CD4⁺ memory cells. Increasingly, there is strong evidence from mouse and human-based scientific studies of a broader diversity in CD4⁺ effector T helper cell subsets. Regulatory T (Treg) cells, have been identified as important negative regulators of adaptive immunity as they limit and suppresses the immune system to control aberrant immune responses to self-antigens; an important mechanism in controlling the development of autoimmune diseases. Follicular helper T (Tfh) cells are another distinct population of effector CD4⁺ T cells that develop from naive T cells post-antigen activation. Tfh cells are specialized in helping B cell humoral immunity as they are uniquely capable of migrating to follicular B cells in secondary lymphoid organs and provide them positive paracrine signals to enable the generation and recall production of high-quality affinity-matured antibodies. Similar to Tregs, Tfh cells also play a role in immunological tolerance as an abnormal expansion of Tfh cell numbers can lead to unrestricted autoreactive antibody production causing severe systemic autoimmune disorders.

The relevance of CD4⁺ T helper cells is highlighted during an HIV infection. HIV is able to subvert the immune system by specifically attacking the CD4⁺ T cells, precisely the cells that could drive the clearance of the virus, but also the cells that drive immunity against all other pathogens encountered during an organism's lifetime.

GAMMA DELTA T CELLS

Gamma delta T cells ($\gamma\delta$ T cells) are T cells that have a distinctive T-cell receptor (TCR) on their surface. Most T cells are $\alpha\beta$ (alpha beta) T cells with TCR composed of two glycoprotein chains called α (alpha) and β (beta) TCR chains. In contrast, gamma delta ($\gamma\delta$) T cells have a TCR that is made up of one γ (gamma) chain and one δ (delta) chain. This group of T cells is usually much less common than $\alpha\beta$ T cells, but are at their highest abundance in the gut mucosa, within a population of lymphocytes known as intraepithelial lymphocytes (IELs). The antigenic molecules that activate gamma delta T cells are still largely unknown. However, $\gamma\delta$ T cells are

peculiar in that they do not seem to require antigen processing and major-histocompatibility-complex (MHC) presentation of peptide epitopes, although some recognize MHC class Ib molecules. Furthermore, $\gamma\delta$ T cells are believed to have a prominent role in recognition of lipid antigens. They are of an invariant nature and may be triggered by alarm signals, such as heat shock proteins (HSP). There also exists a $\gamma\delta$ -T-cell sub-population within the epidermal compartment of the skin of mice. Originally referred to as Thy-1+ dendritic epidermal cells (Thy1+DEC), these cells are more commonly known as dendritic epidermal T cells (DETC). DETCs arise during fetal development and express an invariant and canonical V γ 3 V δ 1 T-cell receptor [using Garman nomenclature].

Gamma Delta T Cells in Innate and Adaptive Immunity

The conditions that lead to responses of gamma delta T cells are not fully understood, and current concepts of them as ‘first line of defense’, ‘regulatory cells’, or ‘bridge between innate and adaptive responses’ only address facets of their complex behaviour. In fact, gamma delta T cells form an entire lymphocyte system that develops under the influence of other leukocytes, in the thymus and in the periphery. When mature they are divided into functionally distinct subsets that obey their own (mostly unknown) rules and that have countless direct and indirect effects on healthy tissues and immune cells, on pathogens and tissues enduring infections and the host responses to them. Like other ‘unconventional’ T cell subsets bearing invariant TCRs, such as CD1d-restricted Natural Killer T cells, gamma delta T cells exhibit several characteristics that place them at the border between the more evolutionarily primitive innate immune system that permits a rapid beneficial response to a variety of foreign agents, and the adaptive immune system, where B and T cells coordinate a slower but highly antigen-specific immune response leading to long-lasting memory against subsequent challenges by the same antigen.

Gamma delta T cells may be considered a component of adaptive immunity in that they rearrange TCR genes to produce junctional diversity and will develop a memory phenotype. However, the various subsets may also be considered part of the innate immunity where a restricted TCR may be used as a pattern recognition receptor. For example, according to this paradigm, large numbers of (human) V γ 9/V δ 2 T cells respond within hours to common molecules produced by microbes, and highly restricted intraepithelial V δ 1 T cells will respond to stressed epithelial cells bearing sentinels of danger. Recent work has shown that human V γ 9/V δ 2 T cells are also capable of phagocytosis, a function previously exclusive to innate myeloid lineage cells such as neutrophils, monocytes and dendritic cells. Clearly, the complexity of gamma delta T cell biology spans definitions of both innate and adaptive immune responses.

GENE FAMILIES IN DIFFERENT SPECIES

Laboratory Mice (*Mus Musculus*)

Mouse V γ Chains

This table summarizes the nomenclature of mouse V γ chains and indicates monoclonal antibodies often used to identify these chains. This system has been best described in strain C57BL/6 and might not apply well to other strains. There are two systems of nomenclature in use (Heilig; Garman), and many writers do not indicate which system they use.

For example, the IMGT (International Immunogenetics Information System) uses the Heilig notation, but does not indicate this fact on its web site. This table refers to variable chain V γ gene segments and to monoclonal antibodies that detect the corresponding V γ protein chains. Note that Adrian Hayday’s proposed nomenclature

is not widely used, leaving considerable confusion in the literature. One advantage and weakness of the Hayday nomenclature is that it is based on the gene order in the B6 genome, but this might not apply to other strains.



Fig. Mouse Vgamma locus for C57BL/6 genome; drawn to scale. Chromosome 13: 1.927 to 1.440 Megabp Heilig notation.

Heilig and Tonegawa's system	Garman's system	"Hayday's system"	Antibodies	Comments
V γ 5	V γ 3	GV1S1	536; 17D1 specific for V γ 5(Heilig)+V δ 1 clonotype	Skin, J γ 1C γ 1
V γ 6	V γ 4	GV2S1	17D1; can detect V γ 6V δ 1 when pretreated with GL3 antibodies	reproductive mucosa; J γ 1C γ 1
V γ 4	V γ 2	GV3S1	UC310A6	lung; J γ 1C γ 1
V γ 7	V γ 5	GV4S1	F2.67 Pereira	most common form in intestinal IEL orthologous to human V γ 1 J γ 1C γ 1
V γ 1	V γ 1.1	GV5S1	2.11 Pereira 1995	peripheral lymphoid tissues; J γ 4C γ 4
V γ 2	V γ 1.2	GV5S2		J γ 1C γ 1
V γ 3	V γ 1.3	GV5S3		J γ 3-pseudoC γ 3

Human Forms

Human V δ 2 T Vells

V γ 9/V δ 2 T cells are unique to humans and primates and represent a minor and unconventional constituent of the leukocyte population in peripheral blood (0.5-5 per cent); yet they are assumed to play an early and essential role in sensing 'danger' by invading pathogens as they expand dramatically in many acute infections and may exceed all other lymphocytes within a few days, *e.g.* in tuberculosis, salmonellosis, ehrlichiosis, brucellosis, tularemia, listeriosis, toxoplasmosis, and malaria. Of note, all V γ 9/V δ 2 T cells recognize the same small microbial compound (*E*)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP), a natural intermediate of the non-mevalonate pathway of isopentenyl pyrophosphate (IPP) biosynthesis. HMB-PP is an essential metabolite in most pathogenic bacteria including *Mycobacterium tuberculosis* and malaria parasites, but is absent from the human host. Bacterial species that lack the non-mevalonate pathway and synthesize IPP via the classical mevalonate pathway instead, such as *Streptococcus*, *Staphylococcus*, and *Borrelia*, are unable to produce HMB-PP and do not specifically activate V γ 9/V δ 2 T cells.

IPP itself is structurally closely related to HMB-PP and ubiquitously present in all living cells (*i.e.* also in human cells), yet its potency *in vitro* is reduced 10,000 fold; whether IPP represents a physiological 'danger' signal of stressed or transformed cells is still unclear. Of pharmacological interest and with bioactivities comparable to that of IPP are synthetic aminobisphosphonates such as zoledronate (Zometa) or pamidronate (Aredia), that are widely used to treat osteoporosis and bone metastases, and incidentally act as V γ 9/V δ 2 T cell receptor agonists. However, increasing evidence suggests that these aminobisphosphonate 'antigens' are not recognised directly by V γ 9/V δ 2 T cells and in fact act indirectly, via their effects on the mevalonate biosynthetic pathway, leading to an accumulation of IPP. Finally, certain alkylated amines have been described to activate V γ 9/V δ 2 T cells *in vitro*, however only at millimolar concentrations, *i.e.* with potencies 10-10fold lower than

those of HMB-PP, thereby questioning their physiological relevance. It is still not clear whether these non-peptidic antigens bind directly to the V γ 9/V δ 2 TCR or if a presenting element exists. There is evidence for a requirement for a species-specific cell-cell contact. However, none of the known antigen-presenting molecules like MHC class I and II or CD1 are required for $\gamma\delta$ T cell activation suggesting the existence of a novel presenting element. Strong support for a direct recognition of non-peptide antigens by the V γ 9/V δ 2 TCR comes from studies which demonstrated that a transfected V γ 9/V δ 2 TCR can confer responsiveness onto a hitherto unresponsive cell; furthermore, antibodies to the $\gamma\delta$ TCR block recognition. Thus, the presence of a functional V γ 9/V δ 2 TCR appears mandatory for a response to non-peptidic antigens although the basis for the huge differences in bioactivity between closely related molecules like HMB-PP and IPP cannot be explained by conventional epitope presentation/recognition models.

These V γ 9V δ 2 T cells can also behave like professional antigen-presenting cells (APC). It seems that human V γ 9V δ 2 T cells are characterized by a specific inflammatory migration programme, including multiple receptors for inflammatory chemokines (CXCR3, CCR1, CCR2 and CCR5). It means that the stimulation with IPP or HMB-PP induces migration to the lymphatic tissues, specifically to the T cell area of lymph nodes. So the stimulation of V γ 9V δ 2 T cells with phosphoantigens results in expression of multiple markers which are associated with APC, like MHC I and II molecules, co-stimulatory molecules (CD80, CD86) and adhesion receptors (CD11a, CD18, CD54). Thus activated V γ 9V δ 2 T cells behave like APCs ($\gamma\delta$ T-APC) and present antigens to $\alpha\beta$ T cells. This leads to turn of naïve CD4+ and CD8+ $\alpha\beta$ T cells into effector cells. The differentiation, induced by $\gamma\delta$ T-APC, most often led to T helper cell response, in the most of cases to pro-inflammatory Th1 response with subsequent production of IFN- γ and TNF- α . But in the case of a low $\gamma\delta$ T-APC: CD4+ ratio it leads to differentiation of some naïve $\alpha\beta$ T cells into Th2 (IL-4) or Th0 (IL-4 plus IFN- γ) cells. Human V γ 9V δ 2 T cells are also cells with excellent antigen cross-presentation activity, a process describing the uptake of exogenous antigen and its routing to the MHC I pathway for induction CD8+ cytotoxic T cells. Thus activated cytotoxic T cells can effectively kill infected or tumor cells. This fact can be used in the immunotherapy of cancer and infectious diseases.

Human non-V δ 2 T Cells

The extensive structural diversity of V δ 1 and V δ 3 TCRs and the existence of V δ 1 clones reactive against MHC, MHC-like, or non-MHC molecules suggest recognition of a highly diverse and heterogeneous set of antigens by non-V δ 2 cells, although cognate interactions between non-V δ 2 TCRs and any of these antigens have not been shown yet. MHC class-I-chain-related gene A (MICA) has also been proposed as an important tumor antigen recognized by V δ 1 T cells. However, the very low affinity of MICA-V δ 1 TCR interactions estimated by surface plasmon resonance analyses raises doubts about the functional relevance of MICA or MHC class-I-chain-related gene B (MICB) recognition by V δ 1 TCRs.

Non-V δ 2 $\gamma\delta$ T cells are expanded in various infectious contexts involving intracellular bacteria (Mycobacteria and Listeria) as well as extracellular bacteria, such as *Borrelia burgdorferi* and viruses (HIV, cytomegalovirus). In most instances, the stimuli that trigger V δ 1 expansion are not derived from pathogens but instead correspond to endogenous gene products presumably upregulated on infection. The antigens recognized by non-V δ 2 T cells expanded in the above infectious contexts have not been characterized, but the fact that V δ 1 T-cell responses are not blocked by monoclonal antibody directed against known classical or non-classical MHC molecules suggests recognition of a new class of conserved stress-induced antigens.

B LYMPHOCYTES AND ANTIBODY PRODUCTION

B Cells are the major cells involved in the creation of antibodies that circulate in blood plasma and lymph, known as humoral immunity. Antibodies (also known as immunoglobulin, Ig), are large Y-shaped proteins used

by the immune system to identify and neutralize foreign objects. In mammals, there are five types of antibody: IgA, IgD, IgE, IgG, and IgM, differing in biological properties; each has evolved to handle different kinds of antigens. Upon activation, B cells produce antibodies, each of which recognize a unique antigen, and neutralizing specific pathogens.

Antigen and antibody binding would cause five different protective mechanism:

- *Agglutination:* Reduces number of infectious units to be dealt with
- *Activation of complement:* Cause inflammation and cell lysis
- *Opsonization:* Coating antigen with antibody enhances phagocytosis
- *Antibody-dependent cell-mediate cytotoxicity:* Antibodies attached to target cell cause destruction by macrophages, eosinophils, and NK cells
- *Neutralization:* Blocks adhesion of bacteria and viruses to mucosa

Like the T cell, B cells express a unique B cell receptor (BCR), in this case, a membrane-bound antibody molecule. All the BCR of any one clone of B cells recognizes and binds to only one particular antigen. A critical difference between B cells and T cells is how each cell “sees” an antigen. T cells recognize their cognate antigen in a processed form – as a peptide in the context of an MHC molecule, whereas B cells recognize antigens in their native form. Once a B cell encounters its cognate (or specific) antigen (and receives additional signals from a helper T cell (predominately Th2 type)), it further differentiates into an effector cell, known as a plasma cell. Plasma cells are short-lived cells (2–3 days) that secrete antibodies. These antibodies bind to antigens, making them easier targets for phagocytes, and trigger the complement cascade. About 10 per cent of plasma cells survive to become long-lived antigen-specific memory B cells. Already primed to produce specific antibodies, these cells can be called upon to respond quickly if the same pathogen re-infects the host, while the host experiences few, if any, symptoms.

ALTERNATIVE ADAPTIVE IMMUNE SYSTEM

Although the classical molecules of the adaptive immune system (*e.g.*, antibodies and T cell receptors) exist only in jawed vertebrates, a distinct lymphocyte-derived molecule has been discovered in primitive jawless vertebrates, such as the lamprey and hagfish. These animals possess a large array of molecules called variable lymphocyte receptors (VLRs for short) that, like the antigen receptors of jawed vertebrates, are produced from only a small number (one or two) of genes. These molecules are believed to bind pathogenic antigens in a similar way to antibodies, and with the same degree of specificity.

IMMUNOLOGICAL MEMORY

When B cells and T cells are activated some become memory B cells and some memory T cells. Throughout the lifetime of an animal these memory cells form a database of effective B and T lymphocytes. Upon interaction with a previously encountered antigen, the appropriate memory cells are selected and activated. In this manner, the second and subsequent exposures to an antigen produce a stronger and faster immune response. This is “adaptive” because the body’s immune system prepares itself for future challenges, but is “maladaptive” of course if the receptors are autoimmune. Immunological memory can be in the form of either *passive* short-term memory or *active* long-term memory.

PASSIVE MEMORY

Passive memory is usually short-term, lasting between a few days and several months. Newborn infants have had no prior exposure to microbes and are particularly vulnerable to infection. Several layers of passive protection are provided by the mother. *In utero*, maternal IgG is transported directly across the placenta, so that,

at birth, human babies have high levels of antibodies, with the same range of antigen specificities as their mother. Breast milk contains antibodies (mainly IgA) that are transferred to the gut of the infant, protecting against bacterial infections, until the newborn can synthesize its own antibodies. This is passive immunity because the fetus does not actually make any memory cells or antibodies: It only borrows them. Short-term passive immunity can also be transferred artificially from one individual to another via antibody-rich serum.

ACTIVE MEMORY

In general, active immunity is long-term and can be acquired by infection followed by B cells and T cells activation, or artificially acquired by vaccines, in a process called immunization.

Immunization

Historically, infectious disease has been the leading cause of death in the human population. Over the last century, two important factors have been developed to combat their spread: sanitation and immunization. Immunization (commonly referred to as vaccination) is the deliberate induction of an immune response, and represents the single most effective manipulation of the immune system that scientists have developed. Immunizations are successful because they utilize the immune system's natural specificity as well as its inducibility. The principle behind immunization is to introduce an antigen, derived from a disease-causing organism, that stimulates the immune system to develop protective immunity against that organism, but that does not itself cause the pathogenic effects of that organism. An antigen (short for *antibody generator*), is defined as any substance that binds to a specific antibody and elicits an adaptive immune response. Most viral vaccines are based on live attenuated viruses, whereas many bacterial vaccines are based on acellular components of microorganisms, including harmless toxin components. Many antigens derived from acellular vaccines do not strongly induce an adaptive response, and most bacterial vaccines require the addition of *adjuvants* that activate the antigen-presenting cells of the innate immune system to enhance immunogenicity.

IMMUNOLOGICAL DIVERSITY

Most large molecules, including virtually all proteins and many polysaccharides, can serve as antigens. The parts of an antigen that interact with an antibody molecule or a lymphocyte receptor, are called epitopes, or antigenic determinants. Most antigens contain a variety of epitopes and can stimulate the production of antibodies, specific T cell responses, or both. A very small proportion (less than 0.01 per cent) of the total lymphocytes are able to bind to a particular antigen, which suggests that only a few cells respond to each antigen. For the adaptive response to “remember” and eliminate a large number of pathogens the immune system must be able to distinguish between many different antigens, and the receptors that recognize antigens must be produced in a huge variety of configurations, in essence one receptor (at least) for each different pathogen that might ever be encountered. Even in the absence of antigen stimulation, a human can produce more than 1 trillion different antibody molecules. Millions of genes would be required to store the genetic information that produces these receptors, but, the entire human genome contains fewer than 25,000 genes.

Myriad receptors are produced through a process known as clonal selection. According to the clonal selection theory, at birth, an animal randomly generates a vast diversity of lymphocytes (each bearing a unique antigen receptor) from information encoded in a small family of genes. To generate each unique antigen receptor, these genes have undergone a process called V(D)J recombination, or *combinatorial diversification*, in which one gene segment recombines with other gene segments to form a single unique gene. This assembly process generates the enormous diversity of receptors and antibodies, before the body ever encounters antigens, and enables the immune system to respond to an almost unlimited diversity of antigens. Throughout an animal's lifetime,

lymphocytes that can react against the antigens an animal actually encounters are selected for action—directed against anything that expresses that antigen. Note that the innate and adaptive portions of the immune system work together, not in spite of each other. The adaptive arm, B, and T cells couldn't function without the innate system's input. T cells are useless without antigen-presenting cells to activate them, and B cells are crippled without T cell help. On the other hand, the innate system would likely be overrun with pathogens without the specialized action of the adaptive immune response.

ADAPTIVE IMMUNITY DURING PREGNANCY

The cornerstone of the immune system is the recognition of “self” versus “non-self”. Therefore, the mechanisms that protect the human fetus (which is considered “non-self”) from attack by the immune system, are particularly interesting. Although no comprehensive explanation has emerged to explain this mysterious, and often repeated, lack of rejection, two classical reasons may explain how the fetus is tolerated. The first is that the fetus occupies a portion of the body protected by a non-immunological barrier, the uterus, which the immune system does not routinely patrol. The second is that the fetus itself may promote local immunosuppression in the mother, perhaps by a process of active nutrient depletion. A more modern explanation for this induction of tolerance is that specific glycoproteins expressed in the uterus during pregnancy suppress the uterine immune response.

During pregnancy in viviparous mammals (all mammals except Monotremes), endogenous retroviruses (ERVs) are activated and produced in high quantities during the implantation of the embryo. They are currently known to possess immunosuppressive properties, suggesting a role in protecting the embryo from its mother's immune system. Also, viral fusion proteins cause the formation of the placental syncytium to limit exchange of migratory cells between the developing embryo and the body of the mother (something an epithelium can't do sufficiently, as certain blood cells specialize to insert themselves between adjacent epithelial cells). The immunodepressive action was the initial normal behaviour of the virus, similar to HIV. The fusion proteins were a way to spread the infection to other cells by simply merging them with the infected one (HIV does this too). It is believed that the ancestors of modern viviparous mammals evolved after an infection by this virus, enabling the fetus to survive the immune system of the mother.

The human genome project found several thousand ERVs classified into 24 families.

IMMUNE NETWORK THEORY

The immune network theory is a theory of how the adaptive immune system works, that has been developed since 1974 mainly by Niels Jerne and Geoffrey W. Hoffmann. The theory states that the immune system is an interacting network of lymphocytes and molecules that have variable (V) regions. These V regions bind not only to things that are foreign to the vertebrate, but also to other V regions within the system. The immune system is therefore seen as a network, with the components connected to each other by V-V interactions. It has been suggested that the phenomena that the theory describes in terms of networks are also explained by clonal selection theory.

The scope of the symmetrical network theory developed by Hoffmann includes the phenomena of low dose and high dose tolerance, first reported for a single antigen by Avron Mitchison, and confirmed by Geoffrey Shellam and Sir Gustav Nossal, the helper and suppressor roles of T cells, the role of non-specific accessory cells in immune responses, and the very important phenomenon called I-J. Jerne was awarded the Nobel Prize for Medicine or Physiology in 1984 partly for his work towards the clonal selection theory, as well as his proposal of the immune network concept. Immune network theory has also inspired a subfield of optimization algorithms similar to artificial neural networks, and unrelated to biological immunology.

THE SYMMETRICAL IMMUNE NETWORK THEORY

Heinz Kohler was involved in early idiotypic network research and was the first to suggest that idiotypic network interactions are symmetrical. developed a detailed immune network theory based on symmetrical stimulatory, inhibitory and killing interactions. It offers a framework for understanding a large number of immunological phenomena based on a small number of postulates. The theory involves roles for B cells that make antibodies, T cells that regulate the production of antibodies by B cells, and non-specific accessory cells (A cells). Antibodies called IgG have two V regions and a molecular weight of 150,000. A central role in the theory is played by specific T cell factors, which have a molecular weight of approximately 50,000, and are postulated in the theory to have only one V region. Hoffmann has proposed that for brevity specific T cell factors should be called tabs. Tabs are able to exert a powerful suppressive effect on the production of IgG antibodies in response to foreign substances (antigens), as was demonstrated rigorously by Takemori and Tada. Hoffmann and Gorczynski have reproduced the Takemori and Tada experiment, confirming the existence of specific T cell factors. In the symmetrical network theory tabs are able to block V regions and also to have a stimulatory role when bound to a tab receptor on A cells. Symmetrical stimulatory interactions follow from the postulate that activation of B cells, T cells and A cells involves cross-linking of receptors. The symmetrical network theory has been developed with the assistance of mathematical modeling. In order to exhibit immune memory to any combination of a large number of different pathogens, the system has a large number of stable steady states. The system is also able to switch between steady states as has been observed experimentally. For example, low or high doses of an antigen can cause the system to switch to a suppressed state for the antigen, while intermediate doses can cause the induction of immunity.

I-J, THE I-J PARADOX, AND A RESOLUTION OF THE I-J PARADOX

The theory accounts for the ability of T cells to have regulatory roles in both helping and suppressing immune responses. In 1976 Murphy et al. and Tada et al. independently reported a phenomenon in mice called I-J. From the perspective of the symmetrical network theory, I-J is one of the most important phenomena in immunology, while for many immunologists who are not familiar with the details of the theory, I-J “does not exist”. In practice I-J is defined by anti-I-J antibodies, that are produced when mice of certain strains are immunized with tissue of certain other strains. I-J was found by these authors to map to within the Major Histocompatibility Complex, but no gene could be found at the site where I-J had been mapped in numerous experiments. The absence of I-J gene(s) within the MHC at the place where I-J had been mapped became known as the “I-J paradox”. This paradox resulted in suppressor T cells and tabs, which both express I-J determinants, falling out of favour, together with the symmetrical network theory, that is based on the existence of tabs. In the meantime however, it has been shown that the I-J paradox can be resolved in the context of the symmetrical network theory. The resolution of the I-J paradox involves a process of mutual selection (or “co-selection”) of suppressor T cells and helper T cells, meaning that (a) those suppressor T cells are selected that have V regions with complementarity to as many helper T cells as possible, and (b) helper T cells are selected not only on the basis of their V regions having some affinity for MHC class II, but also on the basis of the V regions having some affinity for the selected suppressor T cell V regions. The helper T cells and suppressor T cells that are co-selected are then a mutually stabilizing construct, and for a given mouse genome, more than one such mutually stabilizing set can exist. This resolution of the I-J paradox leads to some testable predictions.

RELEVANCE FOR UNDERSTANDING HIV PATHOGENESIS

An immune network model for HIV pathogenesis was published in 1994 postulating that HIV-specific T cells are preferentially infected (Hoffmann, 1994, op cit.). The publication of this paper was followed in 2002

with the publication of a paper entitled “HIV preferentially infects HIV specific CD4+ T cells.” Under the immune network theory, the main cause for progression to AIDS after HIV infection is not the direct killing of infected T helper cells by the virus. Following an infection with HIV that manages to establish itself, there is a complex interaction between the HIV virus, the T helper cells that it infects, and T suppressor cells. These three quasispecies apply selective pressure on one another and co-evolve in such a way that the viral epitopes eventually come to mimic the V regions of the main population of T suppressor cells. Once this happens, anti-HIV antibodies can bind to and kill most of the host’s T suppressor cell population. This results in the dysregulation of the immune system, and eventually to other further anti-self reactions, including against the T helper cell population. At that point, the adaptive immune system is completely compromised and AIDS ensues. Hence in this model, the onset of AIDS is primarily an auto-immune reaction triggered by the cross-reaction of anti-HIV antibodies with T suppressor cells. Once this induced auto-immunity sets in, removing the HIV virus itself (for instance via HAART) would not be sufficient to restore proper immune function. The co-evolution of the quasispecies mentioned above will take a variable time depending on the initial conditions at the time of infection (*i.e.* the epitopes of the first infection and the steady state of the host’s immune cell population), which would explain why there is a variable period, which differs greatly between individual patients, between HIV infection and the onset of AIDS. It also suggests that conventional vaccines are unlikely to be successful, since they would not prevent the auto-immune reaction. In fact such vaccines may do more harm in certain cases, since if the original infection comes from a source with a “mature” infection, those virions will have a high affinity for anti-HIV T helper cells and so increasing the anti-HIV population via vaccination only serves to provide the virus with more easy targets.

AN HIV VACCINE CONCEPT BASED ON IMMUNE NETWORK THEORY

A hypothetical HIV vaccine concept based on immune network theory has been described. The vaccine concept was based on a network theory resolution of the Oudin-Cazenave paradox. This is a phenomenon that makes no sense in the context of clonal selection, without taking idiotypic network interactions into account. The vaccine concept comprised complexes of an anti-anti-HIV antibody and an HIV antigen, and was designed to induce the production of broadly neutralizing anti-HIV antibodies. A suitable anti-anti-HIV antibody envisaged for use in this vaccine is the monoclonal antibody 1F7, which was discovered by Sybille Muller and Heinz Kohler and their colleagues. This monoclonal antibody binds to all of six well characterized broadly neutralizing anti-HIV antibodies.

A MORE GENERAL VACCINE BASED ON IMMUNE NETWORK THEORY

A vaccine concept based on a more recent extension of immune network theory and also based on much more data has been described by Reginald Gorczynski and Geoffrey Hoffmann. The vaccine typically involves three immune systems, A, B and C that can be combined to make an exceptionally strong immune system in a treated vertebrate C. In mouse models the vaccine has been shown to be effective in the prevention of inflammatory bowel disease; the prevention of tumour growth and prevention of metastases in a transplantable breast cancer; and in the treatment of an allergy. The immune system of C is stimulated by a combination of A anti-B (antigen-specific) and B anti-anti-B (antiidiotypic) antibodies. The former stimulate anti-anti-B T cells and the latter stimulate anti-B T cells within C. Mutual selection (“co-selection”) of the anti-B and anti-anti-B T cells takes the system to a new stable steady state in which there are elevated levels of these two populations of T cells. An untreated vertebrate C with self antigens denoted C is believed to have a one-dimensional axis of lymphocytes that is defined by co-selection of anti-C and anti-anti-C lymphocytes. The treated vertebrate C has a two dimensional system of lymphocytes defined by co-selection of both anti-C and anti-anti-C lymphocytes and co-selection of anti-B and anti-anti-B lymphocytes. Experiments indicate that the two-dimensional system is more stable than the one-dimensional system.

MHC RESTRICTION OF V-V INTERACTIONS IN SERUM IGG

IgG from mice with a given set of MHC genes binds rapidly and specifically in an ELISA assay to IgG from mice with the same MHC, but not to IgG from mice with different MHC genes. An idiotypic network model based on co-selection provides an explanation for this phenomenon. In the model IgG molecules are selected to have both anti-anti-(self MHC class II) and anti-anti-anti-(self MHC class II) specificity, and therefore have MHC-restricted self-binding.

EVOLUTION

The adaptive immune system, which has been best-studied in mammals, originated in jawed fish approximately 500 million years ago. Most of the molecules, cells, tissues, and associated mechanisms of this system of defence are found in cartilaginous fishes. Lymphocyte receptors, Ig and TCR, are found in all jawed vertebrates. The most ancient Ig class, IgM, is membrane-bound and then secreted upon stimulation of cartilaginous fish B cells. Another isotype, shark IgW, is related to mammalian IgD. TCRs, both α/β and γ/δ , are found in all animals from gnathostomes to mammals. The organization of gene segments that undergo gene rearrangement differs in cartilaginous fishes, which have a cluster form as compared to the translocon form in bony fish to mammals. Like TCR and Ig, the MHC is found only in jawed vertebrates. Genes involved in antigen processing and presentation, as well as the class I and class II genes, are closely linked within the MHC of almost all studied species. Lymphoid cells can be identified in some pre-vertebrate deuterostomes (*i.e.*, sea urchins). These bind antigen with pattern recognition receptors (PRRs) of the innate immune system. In jawless fishes, two subsets of lymphocytes use variable lymphocyte receptors (VLRs) for antigen binding. Diversity is generated by a cytosine deaminase-mediated rearrangement of LRR-based DNA segments. There is no evidence for the recombination-activating genes (RAGs) that rearrange Ig and TCR gene segments in jawed vertebrates.

The evolution of the AIS, based on Ig, TCR, and MHC molecules, is thought to have arisen from two major evolutionary events: the transfer of the RAG transposon (possibly of viral origin) and two whole genome duplications. Though the molecules of the AIS are well-conserved, they are also rapidly evolving. Yet, a comparative approach finds that many features are quite uniform across taxa. All the major features of the AIS arose early and quickly. Jawless fishes have a different AIS that relies on gene rearrangement to generate diversity but has little else in common with the jawed vertebrate AIS. The innate immune system, which has an important role in AIS activation, is the most important defence system of invertebrates and plants.

TYPES OF ADAPTIVE IMMUNITY

Immunity can be acquired either actively or passively. Immunity is acquired actively when a person is exposed to foreign substances and the immune system responds. Passive immunity is when antibodies are transferred from one host to another. Both actively acquired and passively acquired immunity can be obtained by natural or artificial means.

- *Naturally Acquired Active Immunity*- when a person is naturally exposed to antigens, becomes ill, then recovers.
- *Naturally Acquired Passive Immunity*- involves a natural transfer of antibodies from a mother to her infant. The antibodies cross the woman's placenta to the fetus. Antibodies can also be transferred through breast milk with the secretions of colostrum.
- *Artificially Acquired Active Immunity*- is done by vaccination (introducing dead or weakened antigen to the host's cell).
- *Artificially Acquired Passive Immunity*- This involves the introduction of antibodies rather than antigens to the human body. These antibodies are from an animal or person who is already immune to the disease.

Naturally acquired	Artificially acquired
Active- Antigen enters the body naturally	Active- Antigens are introduced in vaccines.
Passive-Antibodies pass from mother to fetus via placenta or infant via the mother's milk.	Passive- Preformed antibodies in immune serum are introduced by injection.

ANTIGENS

In immunology, an antigen is a molecule capable of inducing an immune response (to produce an antibody) in the host organism. Sometimes antigens are part of the host itself in an autoimmune disease. Antigens are “targeted” by antibodies. Each antibody (immune response) is specifically produced by the immune system to match an antigen after cells in the immune system come into contact with it; this allows a precise identification or matching of the antigen and the initiation of a tailored response. The antibody is said to “match” the antigen in the sense that it can bind to it due to an adaptation performed to a region of the antibody; because of this, many different antibodies are produced, each with specificity to bind a different antigen while sharing the same basic structure. In most cases, an adapted antibody can only react to and bind one specific antigen; in some instances, however, antibodies may cross-react to and bind more than one antigen. Also, an antigen is a molecule that binds to Ag-specific receptors, but cannot necessarily induce an immune response in the body by itself. Antigens are usually peptides (amino acid chains), polysaccharides (chains of monosaccharides/simple sugars) or lipids. In general, saccharides and lipids (as opposed to peptides) qualify as antigens but not as immunogens since they cannot elicit an immune response on their own. Furthermore, for a peptide to induce an immune response (activation of T-cells by antigen-presenting cells) it must be a large enough size, since peptides too small will also not elicit an immune response. The term antigen originally described a structural molecule that binds specifically to an antibody. It was expanded to refer to any molecule or a linear molecular fragment that can be recognized by highly variable antigen receptors (B-cell receptor or T-cell receptor) of the adaptive immune system.

The antigen may originate from within the body (“self-antigen”) or from the external environment (“non-self”). The immune system usually does not react to self-antigens under normal homeostatic conditions due to negative selection of T cells in the thymus and is supposed to identify and attack “non-self” invaders from the outside world or modified/harmful substances present in the body under distressed conditions. Antigen presenting cells present antigens in the form of peptides on histocompatibility molecules. The T cell/T lymphocyte (a subtype of white blood cell), of the adaptive immune system, selectively recognize the antigens. Depending on the antigen and the type of the histocompatibility molecule, different types of T cells will be activated. For T-Cell Receptor (TCR) recognition, the peptide must be processed into small fragments inside the cell and presented by a major histocompatibility complex (MHC). The antigen cannot elicit the immune response without the help of an immunologic adjuvant. Similarly, the adjuvant component of vaccines plays an essential role in the activation of the innate immune system.

An immunogen is an antigen substance (or adduct) that is able to trigger a humoral (innate) or cell-mediated immune response. It first initiates an innate immune response, which then causes the activation of the adaptive immune response. An antigen binds the highly variable immunoreceptor products (B-cell receptor or T-cell receptor) once these have been generated. Immunogens are those antigens, termed immunogenic, capable of inducing an immune response. At the molecular level, an antigen can be characterized by its ability to bind to an antibody’s variable Fab region. Different antibodies have the potential to discriminate among specific epitopes present on the antigen surface. A hapten is a small molecule that changes the structure of an antigenic epitope. In order to induce an immune response, it needs to be attached to a large carrier molecule such as a protein (a complex of peptides). Antigens are usually carried by proteins and polysaccharides, and less frequently, lipids. This includes parts (coats, capsules, cell walls, flagella, fimbriae, and toxins) of bacteria, viruses, and other

microorganisms. Lipids and nucleic acids are antigenic only when combined with proteins and polysaccharides. Non-microbial non-self antigens can include pollen, egg white and proteins from transplanted tissues and organs or on the surface of transfused blood cells. Vaccines are examples of antigens in an immunogenic form, which are intentionally administered to a recipient to induce the memory function of adaptive immune system towards the antigens of the pathogen invading that recipient.

TERMINOLOGY

- **Epitope** – The distinct surface features of an antigen, its *antigenic determinant*. Antigenic molecules, normally “large” biological polymers, usually present surface features that can act as points of interaction for specific antibodies. Any such feature constitutes an epitope. Most antigens have the potential to be bound by multiple antibodies, each of which is specific to one of the antigen’s epitopes. Using the “lock and key” metaphor, the antigen can be seen as a string of keys (epitopes) each of which matches a different lock (antibody). Different antibody idiotypes, each have distinctly formed complementarity determining regions.
- **Allergen** – A substance capable of causing an allergic reaction. The (detrimental) reaction may result after exposure via ingestion, inhalation, injection, or contact with skin.
- **Superantigen** – A class of antigens that cause non-specific activation of T-cells, resulting in polyclonal T cell activation and massive cytokine release.
- **Tolerogen** – A substance that invokes a specific immune non-responsiveness due to its molecular form. If its molecular form is changed, a tolerogen can become an immunogen.
- **Immunoglobulin-binding protein** – Proteins such as Protein A, protein G, and protein L that are capable of binding to antibodies at positions outside of the antigen-binding site. While antigens are the “target” of antibodies, immunoglobulin-binding proteins “attack” antibodies..
- **T-dependent antigen** – Antigens that require the assistance of T cells to induce the formation of specific antibodies.
- **T-independent antigen** – Polysaccharides (usually) that stimulate B cells directly.
- **Immunodominant antigens** – Antigens that dominate (over all others from a pathogen) in their ability to produce an immune response. T cell responses typically are directed against a relatively few immunodominant epitopes, although in some cases (*e.g.*, infection with the malaria pathogen *Plasmodium spp.*) it is dispersed over a relatively large number of parasite antigens.

SOURCES

Antigens can be classified according to their source.

Exogenous Antigens

Exogenous antigens are antigens that have entered the body from the outside, for example by inhalation, ingestion or injection. The immune system’s response to exogenous antigens is often subclinical. By endocytosis or phagocytosis, exogenous antigens are taken into the antigen-presenting cells (APCs) and processed into fragments. APCs then present the fragments to T helper cells (CD4) by the use of class II histocompatibility molecules on their surface. Some T cells are specific for the peptide:MHC complex. They become activated and start to secrete cytokines, substances that activate cytotoxic T lymphocytes (CTL), antibody-secreting B cells, macrophages and other particles. Some antigens start out as exogenous, and later become endogenous (for example, intracellular viruses). Intracellular antigens can be returned to circulation upon the destruction of the infected cell.

Endogenous Antigens

Endogenous antigens are generated within normal cells as a result of normal cell metabolism, or because of viral or intracellular bacterial infection. The fragments are then presented on the cell surface in the complex with MHC class I molecules. If activated cytotoxic CD8 T cells recognize them, the T cells secrete various toxins that cause the lysis or apoptosis of the infected cell. In order to keep the cytotoxic cells from killing cells just for presenting self-proteins, the cytotoxic cells (self-reactive T cells) are deleted as a result of tolerance (negative selection). Endogenous antigens include xenogenic (heterologous), autologous and idiotypic or allogenic (homologous) antigens.

Autoantigens

An autoantigen is usually a normal protein or protein complex (and sometimes DNA or RNA) that is recognized by the immune system of patients suffering from a specific autoimmune disease. Under normal conditions, these antigens should not be the target of the immune system, but in autoimmune diseases, their associated T cells are not deleted and instead attack.

Neoantigens

Neoantigens are those that are entirely absent from the normal human genome. As compared with nonmutated self-antigens, neoantigens are of relevance to tumor control, as the quality of the T cell pool that is available for these antigens is not affected by central T cell tolerance. Technology to systematically analyze T cell reactivity against neoantigens became available only recently.

Viral Antigens

For virus-associated tumors, such as cervical cancer and a subset of head and neck cancers, epitopes derived from viral open reading frames contribute to the pool of neoantigens.

Tumor Antigens

Tumor antigens are those antigens that are presented by MHC class I or MHC class II molecules on the surface of tumor cells. Antigens found only on such cells are called tumor-specific antigens (TSAs) and generally result from a tumor-specific mutation. More common are antigens that are presented by tumor cells and normal cells, called tumor-associated antigens (TAAs). Cytotoxic T lymphocytes that recognize these antigens may be able to destroy tumor cells. Tumor antigens can appear on the surface of the tumor in the form of, for example, a mutated receptor, in which case they are recognized by B cells. For human tumors without a viral etiology, novel peptides (neo-epitopes) are created by tumor-specific DNA alterations.

Process

A large fraction of human tumor mutations are effectively patient-specific. Therefore, neoantigens may also be based on individual tumor genomes. Deep-sequencing technologies can identify mutations within the protein-coding part of the genome (the exome) and predict potential neoantigens. In mice models, for all novel protein sequences, potential MHC-binding peptides were predicted. The resulting set of potential neoantigens was used to assess T cell reactivity. Exome-based analyses were exploited in a clinical setting, to assess reactivity in patients treated by either tumor-infiltrating lymphocyte (TIL) cell therapy or checkpoint blockade. Neoantigen identification was successful for multiple experimental model systems and human malignancies.

The false-negative rate of cancer exome sequencing is low—*i.e.*: the majority of neoantigens occur within exonic sequence with sufficient coverage. However, the vast majority of mutations within expressed genes do not produce neoantigens that are recognized by autologous T cells. As of 2015 mass spectroscopy resolution is insufficient to exclude many false positives from the pool of peptides that may be presented by MHC molecules. Instead, algorithms are used to identify the most likely candidates. These algorithms consider factors such as the likelihood of proteasomal processing, transport into the endoplasmic reticulum, affinity for the relevant MHC class I alleles and gene expression or protein translation levels. The majority of human neoantigens identified in unbiased screens display a high predicted MHC binding affinity. Minor histocompatibility antigens, a conceptually similar antigen class are also correctly identified by MHC binding algorithms. Another potential filter examines whether the mutation is expected to improve MHC binding. The nature of the central TCR-exposed residues of MHC-bound peptides is associated with peptide immunogenicity.

Nativity

A native antigen is an antigen that is not yet processed by an APC to smaller parts. T cells cannot bind native antigens, but require that they be processed by APCs, whereas B cells can be activated by native ones.

ANTIGENIC SPECIFICITY

Antigenic specificity is the ability of the host cells to recognize an antigen specifically as a unique molecular entity and distinguish it from another with exquisite precision. Antigen specificity is due primarily to the side-chain conformations of the antigen. It is measurable and need not be linear or of a rate-limited step or equation.

SUPERANTIGEN

Superantigens (SAGs) are a class of antigens that cause non-specific activation of T-cells resulting in polyclonal T cell activation and massive cytokine release. SAGs are produced by some pathogenic viruses and bacteria most likely as a defence mechanism against the immune system. Compared to a normal antigen-induced T-cell response where 0.0001-0.001 per cent of the body's T-cells are activated, these SAGs are capable of activating up to 20 per cent of the body's T-cells. Furthermore, Anti-CD3 and Anti-CD28 Antibodies (CD28-SuperMAB) have also shown to be highly potent superantigens (and can activate up to 100 per cent of T cells). The large number of activated T-cells generates a massive immune response which is not specific to any particular epitope on the SAG thus undermining one of the fundamental strengths of the adaptive immune system, that is, its ability to target antigens with high specificity. More importantly, the large number of activated T-cells secrete large amounts of cytokines, the most important of which is Interferon gamma. This excess amount of IFN-gamma in turn activates the macrophages. The activated macrophages, in turn, over-produce proinflammatory cytokines such as IL-1, IL-6 and TNF-alpha. TNF-alpha is particularly important as a part of the body's inflammatory response. In normal circumstances it is released locally in low levels and helps the immune system defeat pathogens. However, when it is systemically released in the blood and in high levels (due to mass T-cell activation resulting from the SAG binding), it can cause severe and life-threatening symptoms, including shock and multiple organ failure.

Structure

SAGs are produced intracellularly by bacteria and are released upon infection as extracellular mature toxins.

The sequences of these toxins are relatively conserved among the different subgroups. More important than sequence homology, the 3D structure is very similar among different SAGs resulting in similar functional effects among different groups.

Crystal structures of the enterotoxins reveals that they are compact, ellipsoidal proteins sharing a characteristic two-domain folding pattern comprising an NH₂-terminal β barrel globular domain known as the oligosaccharide/oligonucleotide fold, a long α -helix that diagonally spans the center of the molecule, and a COOH terminal globular domain. The domains have binding regions for the Major Histocompatibility Complex Class II (MHC Class II) and the T cell receptor (TCR), respectively.

Binding

Superantigens bind first to the MHC Class II and then coordinate to the variable alpha or beta chain of T-cell Receptors (TCR)

MHC Class II

SAGs show preference for the HLA-DQ form of the molecule. Binding to the α -chain puts the SAG in the appropriate position to coordinate to the TCR. Less commonly, SAGs attach to the polymorphic MHC class II β -chain in an interaction mediated by a zinc ion coordination complex between three SAG residues and a highly conserved region of the HLA-DR β chain. The use of a zinc ion in binding leads to a higher affinity interaction. Several staphylococcal SAGs are capable of cross-linking MHC molecules by binding to both the α and β chains. This mechanism stimulates cytokine expression and release in antigen presenting cells as well as inducing the production of costimulatory molecules that allow the cell to bind to and activate T cells more effectively.

T-cell Receptor

T-cell binding region of the SAG interacts with the Variable region on the Beta chain of the T-cell Receptor. A given SAG can activate a large proportion of the T-cell population because the human T-cell repertoire comprises only about 50 types of V β elements and some SAGs are capable of binding to multiple types of V β regions. This interaction varies slightly among the different groups of SAGs. Variability among different people in the types of T-cell regions that are prevalent explains why some people respond more strongly to certain SAGs. Group I SAGs contact the V β at the CDR2 and framework region of the molecule. SAGs of Group II interact with the V β region using mechanisms that are conformation-dependent. These interactions are for the most part independent of specific V β amino acid side-chains. Group IV SAGs have been shown to engage all three CDR loops of certain V β forms. The interaction takes place in a cleft between the small and large domains of the SAG and allows the SAG to act as a wedge between the TCR and MHC. This displaces the antigenic peptide away from the TCR and circumvents the normal mechanism for T-cell activation. The biological strength of the SAG (its ability to stimulate) is determined by its affinity for the TCR. SAGs with the highest affinity for the TCR elicit the strongest response. SPMEZ-2 is the most potent SAG discovered to date.

T-cell Signaling

The SAG cross-links the MHC and the TCR inducing a signaling pathway that results in the proliferation of the cell and production of cytokines. This occurs because a cognate antigen activates a T cell not because of its structure *per se*, but because its affinity allows it to bind the TCR for a lengthy enough time period, and the SAG mimics this temporal bonding. Low levels of Zap-70 have been found in T-cells activated by SAGs, indicating that the normal signaling pathway of T-cell activation is impaired. It is hypothesized that Fyn rather than Lck is activated by a tyrosine kinase, leading to the adaptive induction of anergy. Both the protein kinase C pathway and the protein tyrosine kinase pathways are activated, resulting in upregulating production of proinflammatory cytokines. This alternative signaling pathway impairs the calcium/calcineurin and Ras/MAPkinase pathways slightly, but allows for a focused inflammatory response.öé

Direct Effects

SAg stimulation of antigen presenting cells and T-cells elicits a response that is mainly inflammatory, focused on the action of Th1 T-helper cells. Some of the major products are IL-1, IL-2, IL-6, TNF- α , gamma interferon (IFN- γ), macrophage inflammatory protein 1 α (MIP-1 α), MIP-1 β , and monocyte chemoattractant protein 1 (MCP-1). This excessive uncoordinated release of cytokines, (especially TNF- α), overloads the body and results in rashes, fever, and can lead to multi-organ failure, coma and death. Deletion or anergy of activated T-cells follows infection. This results from production of IL-4 and IL-10 from prolonged exposure to the toxin. The IL-4 and IL-10 downregulate production of IFN-gamma, MHC Class II, and costimulatory molecules on the surface of APCs. These effects produce memory cells that are unresponsive to antigen stimulation. One mechanism by which this is possible involves cytokine-mediated suppression of T-cells. MHC crosslinking also activates a signaling pathway that suppresses hematopoiesis and upregulates Fas-mediated apoptosis. IFN- α is another product of prolonged SAg exposure. This cytokine is closely linked with induction of autoimmunity, and the autoimmune disease Kawasaki Disease is known to be caused by SAg infection. SAg activation in T-cells leads to production of CD40 ligand which activates isotype switching in B cells to IgG and IgM and IgE. To summarize, the T-cells are stimulated and produce excess amounts of cytokine resulting in cytokine-mediated suppression of T-cells and deletion of the activated cells as the body returns to homeostasis. The toxic effects of the microbe and SAg also damage tissue and organ systems, a condition known as Toxic Shock Syndrome. If the initial inflammation is survived, the host cells become anergic or are deleted, resulting in a severely compromised immune system.

Superantigenicity Independent (Indirect) Effects

Apart from their mitogenic activity, SAGs are able to cause symptoms that are characteristic of infection.

One such effect is emesis. This effect is felt in cases of food poisoning, when SAg-producing bacteria release the toxin, which is highly resistant to heat. There is a distinct region of the molecule that is active in inducing gastrointestinal toxicity. This activity is also highly potent, and quantities as small as 20-35 μ g of SAg are able to induce vomiting. SAGs are able to stimulate recruitment of neutrophils to the site of infection in a way that is independent of T-cell stimulation. This effect is due to the ability of SAGs to activate monocytic cells, stimulating the release of the cytokine TNF- α , leading to increased expression of adhesion molecules that recruit leukocytes to infected regions. This causes inflammation in the lungs, intestinal tissue, and any place that the bacteria have colonized. While small amounts of inflammation are natural and helpful, excessive inflammation can lead to tissue destruction. One of the more dangerous indirect effects of SAg infection concerns the ability of SAGs to augment the effects of endotoxins in the body. This is accomplished by reducing the threshold for endotoxicity. Schlievert demonstrated that, when administered conjunctively, the effects of SAg and endotoxin are magnified as much as 50,000 times. This could be due to the reduced immune system efficiency induced by SAg infection. Aside from the synergistic relationship between endotoxin and SAg, the “double hit” effect of the activity of the endotoxin and the SAg result in effects more deleterious than those seen in a typical bacterial infection. This also implicates SAGs in the progression of sepsis in patients with bacterial infections.

Diseases Associated with Superantigen Production

- Diabetes mellitus
- Eczema
- Guttate psoriasis
- Kawasaki Disease
- Nasal polyps

- Rheumatic fever
- Rheumatoid arthritis
- Scarlet fever
- Toxic Shock Syndrome
- Infective Endocarditis

Treatment

The primary goals of medical treatment are to hemodynamically stabilize the patient and to eliminate the microbe that is producing the SAg. This is accomplished through the use of vasopressors, fluid resuscitation and antibiotics. The body naturally produces antibodies to some SAg, and this effect can be augmented by stimulating B-cell production of these antibodies. Immunoglobulin pools are able to neutralize specific antibodies and prevent T-cell activation. Synthetic antibodies and peptides have been created to mimic SAg-binding regions on the MHC class II, blocking the interaction and preventing T cell activation. Immunosuppressants are also employed to prevent T-cell activation and the release of cytokines. Corticosteroids are used to reduce inflammatory effects.

Evolution of Superantigen Production

SAg production effectively corrupts the immune response, allowing the microbe secreting the SAg to be carried and transmitted unchecked. One mechanism by which this is done is through inducing anergy of the T-cells to antigens and SAg. Lussow and MacDonald demonstrated this by systematically exposing animals to a streptococcal antigen. They found that exposure to other antigens after SAg infection failed to elicit an immune response. In another experiment, Watson and Lee discovered that memory T-cells created by normal antigen stimulation were anergic to SAg stimulation and that memory T-cells created after a SAg infection were anergic to all antigen stimulation. The mechanism by which this occurred was undetermined. The genes that regulate SAg expression also regulate mechanisms of immune evasion such as M protein and Bacterial capsule expression, supporting the hypothesis that SAg production evolved primarily as a mechanism of immune evasion. When the structure of individual SAg domains has been compared to other immunoglobulin-binding streptococcal proteins (such as those toxins produced by *E. coli*) it was found that the domains separately resemble members of these families. This homology suggests that the SAg evolved through the recombination of two smaller β -strand motifs.

Endogenous SAg

Minor lymphocyte stimulating (Mls) exotoxins were originally discovered in the thymic stromal cells of mice. These toxins are encoded by SAg genes that were incorporated into the mouse genome from the mouse mammary tumour virus (MMTV). The presence of these genes in the mouse genome allows the mouse to express the antigen in the thymus as a means of negatively selecting for lymphocytes with a variable Beta region that is susceptible to stimulation by the viral SAg. The result is that these mice are immune to infection by the virus later in life.

Similar endogenous SAg-dependent selection has yet to be identified in the human genome, but endogenous SAg have been discovered and are suspected of playing an integral role in viral infection. Infection by the Epstein-Barr virus, for example, is known to cause production of a SAg in infected cells, yet no gene for the toxin has been found on the genome of the virus. The virus manipulates the infected cell to express its own SAg genes, and this helps it to evade the host immune system. Similar results have been found with rabies, cytomegalovirus, and HIV.

ALLERGEN

An allergen is a type of antigen that produces an abnormally vigorous immune response in which the immune system fights off a perceived threat that would otherwise be harmless to the body. Such reactions are called allergies. In technical terms, an allergen is an antigen capable of stimulating a type-I hypersensitivity reaction in atopic individuals through Immunoglobulin E (IgE) responses. Most humans mount significant Immunoglobulin E responses only as a defence against parasitic infections. However, some individuals may respond to many common environmental antigens. This hereditary predisposition is called atopy. In atopic individuals, non-parasitic antigens stimulate inappropriate IgE production, leading to type I hypersensitivity. Sensitivities vary widely from one person (or other animal) to another. A very broad range of substances can be allergens to sensitive individuals.

Types of Allergens

Allergens can be found in a variety of sources, such as dust mite excretion, pollen, pet dander, or even royal jelly. Food allergies are not as common as food sensitivity, but some foods such as peanuts (a legume), nuts, seafood and shellfish are the cause of serious allergies in many people. Officially, the United States Food and Drug Administration does recognize eight foods as being common for allergic reactions in a large segment of the sensitive population. These include peanuts, tree nuts, eggs, milk, shellfish, fish, wheat and their derivatives, and soy and their derivatives, as well as sulfites (chemical-based, often found in flavors and colors in foods) at 10ppm and over. Other countries, in view of the differences in the genetic profiles of their citizens and different levels of exposure to specific foods due to different dietary habits, the “official” allergen list will change. Canada recognizes all eight of the allergens recognized by the US, and also recognizes sesame seeds, and mustard.

The European Union additionally recognizes celery. Another allergen is urushiol, a resin produced by poison ivy and poison oak, which causes the skin rash condition known as urushiol-induced contact dermatitis by changing a skin cell’s configuration so that it is no longer recognized by the immune system as part of the body. Various trees and wood products such as paper, cardboard, MDF etc. can also cause mild to severe allergy symptoms through touch or inhalation of sawdust such as asthma and skin rash. An allergic reaction can be caused by any form of direct contact with the allergen—consuming food or drink one is sensitive to (ingestion), breathing in pollen, perfume or pet dander (inhalation), or brushing a body part against an allergy-causing plant (direct contact). Other common causes of serious allergy are wasp, fire ant and bee stings, penicillin, and latex. An extremely serious form of an allergic reaction is called anaphylaxis. One form of treatment is the administration of sterile epinephrine to the person experiencing anaphylaxis, which suppresses the body’s overreaction to the allergen, and allows for the patient to be transported to a medical facility.

Common Allergens

In addition to foreign proteins found in foreign serum (from blood transfusions) and vaccines, common allergens include:

- Animal products
 - (a) Fel d 1 (Allergy to cats)
 - (b) Ur and dander
 - (d) Cockroach calyx
 - (e) Wool
 - (f) Dust mite excretion
- Drugs

- (a) Penicillin
- (b) Sulfonamides
- (c) Salicylates (also found naturally in numerous fruits)
- Foods
 - (a) Celery and celeriac
 - (b) Corn or maize
 - (c) Eggs (typically albumen, the white)
 - (d) Fruit
 - Pumpkin, egg-plant
 - (e) Legumes
 - Beans
 - Peas
 - Peanuts
 - Soybeans
 - (f) Milk
 - (g) Seafood
 - (h) Sesame
 - (i) Soy
 - (j) Tree nuts
 - Pecans
 - Almonds
 - (k) Wheat
- Insect stings
 - (a) Bee sting venom
 - (b) Wasp sting venom
 - (c) Mosquito stings
- Mold spores
- Top 5 allergens discovered in patch tests in 2005–06:
 - (a) Nickel sulfate (19.0 per cent)
 - (b) Balsam of Peru (11.9 per cent)
 - (c) Fragrance mix I (11.5 per cent)
 - (d) Quaternium-15 (10.3 per cent), and
 - (e) Neomycin (10.0 per cent).
- Metals
 - (a) Nickel
 - (b) Chromium
 - (c) Cadmium
- Other
 - (a) Latex
 - (b) Wood
- Plant pollens (hay fever)
 - (a) Grass — ryegrass, timothy-grass
 - (b) Weeds — ragweed, plantago, nettle, *Artemisia vulgaris*, *Chenopodium album*, sorrel
 - (c) Trees — birch, alder, hazel, hornbeam, Aesculus, willow, poplar, Platanus, Tilia, Olea, Ashe juniper, *Alstonia scholaris*

Seasonal Allergy

Seasonal allergy symptoms are commonly experienced during specific parts of the year, usually during spring, summer or fall when certain trees or grasses pollinate. This depends on the kind of tree or grass. For instance, some trees such as oak, elm, and maple pollinate in the spring, while grasses such as Bermuda, timothy and orchard pollinate in the summer. Grass allergy is generally linked to hay fever because their symptoms and causes are somehow similar to each other. Symptoms include rhinitis, which causes sneezing and a runny nose, as well as allergic conjunctivitis, which includes watering and itchy eyes. Also an initial tickle on the roof of the mouth or in the back of the throat may be experienced. Also, depending on the season, the symptoms may be more severe and people may experience coughing, wheezing, and irritability. A few people even become depressed, lose their appetite, or have problems sleeping. Moreover, since the sinuses may also become congested, some people experience headaches.

If both parents suffered from allergies in the past, there is a 66 per cent chance for the individual to suffer from seasonal allergies, and the risk lowers to 60 per cent if just one parent had suffered from allergies. The immune system also has strong influence on seasonal allergies, since it reacts differently to diverse allergens like pollen. When an allergen enters the body of an individual that is predisposed to allergies, it triggers an immune reaction and the production of antibodies. These allergen antibodies migrate to mast cells lining the nose, eyes and lungs. When an allergen drifts into the nose more than once, mast cells release a slew of chemicals or histamines that irritate and inflame the moist membranes lining the nose and produce the symptoms of an allergic reaction: scratchy throat, itching, sneezing and watery eyes. Some symptoms that differentiate allergies from a cold include:

- No fever.
- Mucous secretions are runny and clear.
- Sneezes occurring in rapid and several sequences.
- Itchy throat, ears and nose.
- These symptoms usually last longer than 7–10 days.

Among seasonal allergies, there are some allergens that fuse together and produce a new type of allergy. For instance, grass pollen allergens cross-react with food allergy proteins in vegetables such as onion, lettuce, carrots, celery and corn. Besides, the cousins of birch pollen allergens, like apples, grapes, peaches, celery and apricots, produce severe itching in the ears and throat. The cypress pollen allergy brings a cross reactivity between diverse species like olive, privet, ash and Russian olive tree pollen allergens. In some rural areas there is another form of seasonal grass allergy, combining airborne particles of pollen mixed with mold. Recent research has suggested that humans might develop allergies as a defence to fight off parasites. According to Yale University Immunologist Dr Ruslan Medzhitov, protease allergens cleave the same sensor proteins that evolved to detect proteases produced by the parasitic worms. Additionally, a new report on seasonal allergies called “Extreme allergies and Global Warming”, have found that many allergy triggers are worsening due to climate change. 16 states in the United States were named as “Allergen Hotspots” for large increases in allergenic tree pollen if global warming pollution keeps increasing. Therefore, researchers on this report claimed that global warming is bad news for millions of asthmatics in the United States whose asthma attacks are triggered by seasonal allergies. Indeed, seasonal allergies are one of the main triggers for asthma, along with colds or flu, cigarette smoke and exercise. In Canada, for example, up to 75 per cent of asthmatics also have seasonal allergies.

Seasonal Allergy Diagnosis

Based on the symptoms seen on the patient, the answers given in terms of symptom evaluation and a physical exam, doctors can make a diagnosis to identify if the patient has a seasonal allergy. After performing the

diagnosis, the doctor is able to tell the main cause of the allergic reaction and recommend the treatment to follow. 2 tests have to be done in order to determine the cause: a blood test and a skin test. Allergists do skin tests in one of two ways: either dropping some purified liquid of the allergen onto the skin and pricking the area with a small needle; or injecting a small amount of allergen under the skin. Alternative tools are available to identify seasonal allergies, such as laboratory tests, imaging tests and nasal endoscopy. In the laboratory tests, the doctor will take a nasal smear and it will be examined microscopically for factors that may indicate a cause: increased numbers of eosinophils (white blood cells), which indicates an allergic condition. If there is a high count of eosinophils, an allergic condition might be present. Another laboratory test is the blood test for IgE (immunoglobulin production), such as the radioallergosorbent test (RAST) or the more recent enzyme allergosorbent tests (EAST), implemented to detect high levels of allergen-specific IgE in response to particular allergens. Although blood tests are less accurate than the skin tests, they can be performed on patients unable to undergo skin testing. Imaging tests can be useful to detect sinusitis in people suffering from chronic rhinitis, and they can work when other test results are ambiguous. There is also nasal endoscopy, wherein a tube is inserted through the nose with a small camera to view the passageways and examine any irregularities in the nose structure. Endoscopy can be used for some cases of chronic or unresponsive seasonal rhinitis.

Fungal Allergens

In 1952 basidiospores were described as being possible airborne allergens and were linked to asthma in 1969. Basidiospores are the dominant airborne fungal allergens. Fungal allergies are associated with seasonal asthma. They are considered to be a major source of airborne allergens. The basidiospore family include mushrooms, rusts, smuts, brackets, and puffballs. The airborne spores from mushrooms reach levels comparable to those of mold and pollens. The levels of mushroom respiratory allergy are as high as 30 percent of those with allergic disorder, but it is believed to be less than 1 percent of food allergies. Heavy rainfall (which increases fungal spore release) is associated with increased hospital admissions of children with asthma. A study in New Zealand found that 22 percent of patients with respiratory allergic disorders tested positive for basidiospores allergies. Mushroom spore allergies can cause either immediate allergic symptomatology or delayed allergic reactions. Those with asthma are more likely to have immediate allergic reactions and those with allergic rhinitis are more likely to have delayed allergic responses. A study found that 27 percent of patients were allergic to basidiomycete mycelia extracts and 32 percent were allergic to basidiospore extracts, thus demonstrating the high incidence of fungal sensitisation in individuals with suspected allergies. It has been found that of basidiomycete cap, mycelia, and spore extracts that spore extracts are the most reliable extract for diagnosing basidiomycete allergy. In Canada, 8 per cent of children attending allergy clinics were found to be allergic to *Ganoderma*, a basidiospore. *Pleurotus ostreatus*, *cladosporium*, and *Calvatia cyathiformis* are significant airborne spores. Other significant fungal allergens include *aspergillus* and *alternaria*-penicillin families. In India *Fomes pectinatus* is a predominant air-borne allergen affecting up to 22 percent of patients with respiratory allergies. Some fungal air-borne allergens such as *Coprinus comatus* are associated with worsening of eczematous skin lesions. Children who are born during autumn months (during fungal spore season) are more likely to develop asthmatic symptoms later in life.

Treatment

Treatment includes over-the-counter medications, antihistamines, nasal decongestants, allergy shots, and alternative medicine. In the case of nasal symptoms, antihistamines are normally the first option. They may be taken together with pseudoephedrine to help relieve a stuffy nose and they can stop the itching and sneezing. Some over-the-counter options are Benadryl and Tavist. However, these antihistamines may cause extreme drowsiness, therefore, people are advised to not operate heavy machinery or drive while taking this kind of

medication. Other side effects include dry mouth, blurred vision, constipation, difficulty with urination, confusion, and light-headedness. There is also a newer second generation of antihistamines that are generally classified as the “non-sedating antihistamines” or anti-drowsy, which include cetirizine, loratadine, and fexofenadine. An example of nasal decongestants is pseudoephedrine and its side-effects include insomnia, restlessness, and difficulty urinating. Some other nasal sprays are available by prescription, including Azelastine and Ipratropium. Some of their side-effects include drowsiness. For eye symptoms, it is important to first bath the eyes with plain eyewashes to reduce the irritation. People should not wear contact lenses during episodes of conjunctivitis.

Allergy immunotherapy (AIT) treatment involves administering doses of allergens to accustom the body to induce specific long-term tolerance. Allergy immunotherapy can be administered orally (as sublingual tablets or sublingual drops), or by injections under the skin (subcutaneous). Discovered by Leonard Noon and John Freeman in 1911, allergy immunotherapy represents the only causative treatment for respiratory allergies. Immunotherapy contains a small amount of the substance that triggers the allergic reactions and it should start after the pollen season to get prepared for the next season. Natural remedies are another option that patients look to for relief. One of the most popular recently is the European herb butterbur (*Petasites hybridus*). The British Medical Journal published a study in which Swiss researchers proved how one tablet of butterbur four times daily was as effective as an antihistamine in controlling symptoms of hay fever. On a second study, a group of British researchers gave their approval to butterbur’s effectiveness in battling symptoms of grass allergy. Other herbal supplements that function as remedies include freeze-dried nettles (*Urtica*) and a tonic made from the goldenseal herb (*Hydrastis candelensis*), which doctors recommend in addition to saline nasal spray (another natural remedy).

HAPTEN

Haptens are minute molecules that elicit an immune response only when attached to a large carrier such as a protein; the carrier may be one that also does not elicit an immune response by itself. (In general, only large molecules, infectious agents, or insoluble foreign matter can elicit an immune response in the body.) Once the body has generated antibodies to a hapten-carrier adduct, the small-molecule hapten may also be able to bind to the antibody, but it will usually not initiate an immune response; usually only the hapten-carrier adduct can do this. Sometimes the small-molecule hapten can even block immune response to the hapten-carrier adduct by preventing the adduct from binding to the antibody, a process called *hapten inhibition*. The mechanisms of absence of immune response may vary and involve complex immunological mechanisms, but can include absent or insufficient co-stimulatory signals from antigen-presenting cells. Haptens have been used to study allergic contact dermatitis (ACD) and the mechanisms of inflammatory bowel disease (IBD) to induce autoimmune-like responses. The concept of haptens emerged from the work of Karl Landsteiner who also pioneered the use of synthetic haptens to study immunochemical phenomena.

Examples of Haptens

The first researched haptens were aniline and its carboxyl derivatives (o-, m-, and p-aminobenzoic acid).

A well-known example of a hapten is urushiol, which is the toxin found in poison ivy. When absorbed through the skin from a poison ivy plant, urushiol undergoes oxidation in the skin cells to generate the actual hapten, a reactive quinone-type molecule, which then reacts with skin proteins to form hapten adducts. Usually, the first exposure causes only sensitization, in which there is a proliferation of effector T-cells. After a subsequent, second exposure, the proliferated T-cells can become activated, generating an immune reaction that produces typical blisters of a poison ivy exposure. Some haptens can induce autoimmune disease. An example is hydralazine, a blood pressure-lowering drug that occasionally can produce drug-induced lupus erythematosus in certain individuals. This also appears to be the mechanism by which the anaesthetic gas halothane can cause

a life-threatening hepatitis, as well as the mechanism by which penicillin-class drugs cause autoimmune hemolytic anemia. Other haptens that are commonly used in molecular biology applications include fluorescein, biotin, digoxigenin, and dinitrophenol. Lastly, nickel allergy is caused by nickel metal ions penetrating the skin and binding to skin proteins.

Hapten Conjugation

Due to their nature and properties, hapten-carrier adducts have been essential in immunology. They have been used to evaluate the properties of specific epitopes and antibodies. They are important in the purification and production of monoclonal antibodies. They are also vital in the development of sensitive quantitative and qualitative immunoassays. However, to achieve the best and most desirable results, many factors are needed to be taken into the design of hapten conjugates. These include the method of hapten conjugation, the type of carrier used and the hapten density. Variations in these factors could lead to different strengths of immune response towards the newly formed antigenic determinant.

Carrier Selection

A suitable carrier must be chosen. In general, these carrier proteins should be immunogenic and contain enough amino acid residues in the reactive side chains to conjugate with the haptens. Depending on the haptens being used, other factors in considering the carrier proteins could include their *in vivo* toxicity, commercial availability and cost.

The most common carriers include serum globulin, albumins, ovalbumin and many others. Although proteins are mostly employed for hapten conjugation, synthetic polypeptides such as Poly-L-glutamic acid, polysaccharides and liposomes could also be used.

Methods of Hapten Conjugation

While selecting a suitable method for hapten conjugation, functional groups on the hapten and its carrier must be identified. Depending on the groups present, one of the two main strategies could be employed:

- *Spontaneous Chemical Reaction:* Used when hapten is a chemical reactive molecule such as anhydrides and isocyanates. This method of conjugation is spontaneous and no cross-linking agents are needed.
- *Intermediary Molecules Cross-linkage:* This method mainly applies to nonreactive haptens. Agents with at least two chemically reactive groups such as carbodiimide or glutaraldehyde are to aid the conjugation of haptens to their carriers. The extent of cross-linkage is dependent upon the hapten/carrier to coupling agent ratio, hapten/carrier concentration and the temperature, pH of the environment.
 - (a) *Carbodiimide:* A group of compounds with a general formula of $R-N=C=N-R_2$, where R and R_2 are either aliphatic (*i.e.*, diethylcarbodiimide) or aromatic (*i.e.*, diphenylcarbodiimide). Conjugation using a carbodiimide requires the presence of α or ϵ -amino and a carboxyl group. The amino group usually comes from the lysyl residue of the carrier protein while the carboxyl group comes from the hapten. The exact mechanism for this reaction is still unknown. However, two pathways are proposed. The first postulates that an intermediate that can react with an amine is formed. The second stating that a rearrangement of an acyl urea, the main side product of the reaction at high temperature, has occurred.
 - (α) *Glutaraldehyde:* This method works by the reaction between glutaraldehyde with amine groups to form Schiff bases or Michael-type double bond addition products. The yield of conjugates can be controlled by varying the pH of the reaction. Higher pH would give rise to more Schiff

base intermediates and subsequently lead to the increase in hapten conjugates' number and size. Overall, cross-linkage involving glutaraldehyde is very stable. However, immunized animals tend to recognise glutaraldehyde's cross-linking bridges as epitopes.

- *High Performance Capillary Electrophoresis*: High Performance Capillary Electrophoresis (HPCE) is an alternative method in optimizing hapten-protein conjugation. HPCE is predominantly used in separating carbohydrates with a very high separation capacity. There are numerous advantages to using HPCE as a technique to investigate certain conjugates such as only requiring minute sample sizes (nl). In addition, the sample used does not need to be pure and no type of radiolabeling is needed. A great benefit to this method of hapten conjugation is that there is automated analysis of sample and the testing of sample interactions can be determined in free solution. This method of hapten-protein conjugation is exceptionally effective with conjugates of low epitope densities, where it is otherwise very challenging by the use of other methods to determine their electrical or ionic mobility.

Hapten Inhibition

Hapten inhibition or “semi-hapten” is the inhibition of a type III hypersensitivity response. In inhibition, free hapten molecules bind with antibodies towards that molecule without causing the immune response, leaving fewer antibodies left to bind to the immunogenic hapten-protein adduct. An example of a *hapten inhibitor* is dextran 1, which is a small fraction (1 kilodalton) of the entire dextran complex, which is enough to bind anti-dextran antibodies, but insufficient to result in the formation of immune complexes and resultant immune responses.

EPITOPE

An epitope, also known as antigenic determinant, is the part of an antigen that is recognized by the immune system, specifically by antibodies, B cells, or T cells. For example, the epitope is the specific piece of the antigen to which an antibody binds. The part of an antibody that binds to the epitope is called a paratope. Although epitopes are usually non-self proteins, sequences derived from the host that can be recognized (as in the case of autoimmune diseases) are also epitopes. The epitopes of protein antigens are divided into two categories, conformational epitopes and linear epitopes, based on their structure and interaction with the paratope. A conformational epitope is composed of discontinuous sections of the antigen's amino acid sequence. These epitopes interact with the paratope based on the 3-D surface features and shape or tertiary structure of the antigen. The proportion of epitopes that are conformational is unknown. By contrast, linear epitopes interact with the paratope based on their primary structure. A linear epitope is formed by a continuous sequence of amino acids from the antigen.

Function

T Cell Epitopes

T cell epitopes are presented on the surface of an antigen-presenting cell, where they are bound to MHC molecules. In humans, professional antigen-presenting cells are specialized to present MHC class II peptides, whereas most nucleated somatic cells present MHC class I peptides. T cell epitopes presented by MHC class I molecules are typically peptides between 8 and 11 amino acids in length, whereas MHC class II molecules present longer peptides, 13-17 amino acids in length, and non-classical MHC molecules also present non-peptidic epitopes such as glycolipids.

Cross-activity

Epitopes are sometimes cross-reactive. This property is exploited by the immune system in regulation by anti-idiotypic antibodies (originally proposed by Nobel laureate Niels Kaj Jerne). If an antibody binds to an antigen's epitope, the paratope could become the epitope for another antibody that will then bind to it. If this second antibody is of IgM class, its binding can upregulate the immune response; if the second antibody is of IgG class, its binding can downregulate the immune response.

Epitope Mapping

Epitopes can be mapped using protein microarrays, and with the ELISPOT or ELISA techniques. Another technique involves high-throughput mutagenesis, an epitope mapping strategy developed to improve rapid mapping of conformational epitopes on structurally complex proteins. MHC class I and II epitopes can be reliably predicted by computational means alone, although not all in-silico T cell epitope prediction algorithms are equivalent in their accuracy.

Epitope Tags

Epitopes are often used in proteomics and the study of other gene products. Using recombinant DNA techniques genetic sequences coding for epitopes that are recognized by common antibodies can be fused to the gene. Following synthesis, the resulting epitope tag allows the antibody to find the protein or other gene product enabling lab techniques for localisation, purification, and further molecular characterization. Common epitopes used for this purpose are Myc-tag, HA-tag, FLAG-tag, GST-tag, 6xHis and OLLAS. Peptides can also be bound by proteins that form covalent bonds to the peptide, allowing irreversible immobilisation. These strategies have also been successfully applied to the development of "epitope-focused" vaccine design.

Neoantigenic Determinant

A neoantigenic determinant is an epitope on a neoantigen, which is a newly formed antigen that has not been previously recognized by the immune system. Neoantigens are often associated with tumor antigens and are found in oncogenic cells. Neoantigens and, by extension, neoantigenic determinants can be formed when a protein undergoes further modification within a biochemical pathway such as glycosylation, phosphorylation or proteolysis. This, by altering the structure of the protein, can produce new epitopes that are called neoantigenic determinants as they give rise to new antigenic determinants. Recognition requires separate, specific antibodies.

Linear Epitope

A linear or a sequential epitope is an epitope that is recognized by antibodies by its linear sequence of amino acids, or primary structure. In contrast, most antibodies recognize a conformational epitope that has a specific three-dimensional shape and its protein structure. An antigen is any substance that the immune system can recognize as being foreign and which provokes an immune response. Since antigens are usually proteins that are too large to bind as a whole to any receptor, only specific segments that form the antigen bind with a specific antibody. Such segments are called epitopes. Likewise, it is only paratope of the antibody that comes in contact with the epitope. Proteins are composed of repeating nitrogen-containing subunits called amino acids. The linear sequence of amino acids that compose a protein is called its primary structure, which typically does not present as simple line of sequential proteins (much like a knot, rather than a straight string). But, when an antigen is broken down in a lysosome, it yields small peptides, which can be recognized through the amino acids that lie continuously in a line, and hence are called *linear epitopes*.

Significance

While performing molecular assays involving use of antibodies such as in the Western blot, immunohistochemistry, and ELISA, one should carefully choose antibodies that recognize linear or conformational epitopes. For instance, if a protein sample is boiled, treated with beta-mercaptoethanol, and run in SDS-PAGE for the Western blot, the proteins are essentially denatured and therefore cannot assume their natural three-dimensional conformations. Therefore, antibodies that recognize linear epitopes instead of conformational epitopes are chosen for immunodetection. In contrast, in immunohistochemistry where protein structure is preserved, antibodies are preferred that recognize conformational epitopes.

Conformational Epitope

A conformational epitope is a sequence of sub-units (usually amino acids) composing an antigen that come in direct contact with a receptor of the immune system. An antigen is any substance that the immune system can recognize as foreign. Antigens are usually proteins that are too large to bind as a whole to any receptor so only specific segments, that form the antigen, bind with a specific receptor. Such segments are called epitopes. Likewise, it is only paratope of the receptor that comes in contact with the epitope. Proteins are composed of repeating nitrogen-containing subunits called amino acids that in nature do not exist as straight chains called primary structure, but as folded whorls with complex loops. The latter is known as the tertiary structure of a protein. So, whenever a receptor interacts with an undigested antigen, the surface amino acids that come in contact may not be continuous with each other if the protein is unwound. Such discontinuous amino acids that come together in three-dimensional conformation and interact with the receptor's paratope are called conformational epitopes. In contrast, if the antigen is digested, small segments called peptides are formed, which bind with major histocompatibility complex molecules, and then later with T cell receptors through amino acids that are continuous in a line. These are known as linear epitopes.

MIMOTOPE

A mimotope is a macromolecule, often a peptide, which mimics the structure of an epitope. Because of this property it causes an antibody response similar to the one elicited by the epitope. An antibody for a given epitope antigen will recognize a mimotope which mimics that epitope. Mimotopes are commonly obtained from phage display libraries through biopanning. Vaccines utilizing mimotopes are being developed. When the term mimotope was first coined by Mario Geysen in 1986, it was used to describe peptides mimicking epitope. However, this concept has been extended to refer peptide mimic of all types of binding sites. As the mimic of binding site, mimotope analysis has been widely used in mapping epitopes, identifying drug target and inferring protein interaction networks. Furthermore, mimotope has also shown its potential in the development of new diagnostics, therapeutics and vaccines. In addition, special affinities mediated by mimotopes to various semiconductors and other materials have shown very encouraging promise in new material and new energy studies.

Gathering information on mimotopes into a special database therefore deserves. In 2010, the MimoDB database version 1.0 was released. It had 10716 peptides grouped into 1229 sets. These peptides were extracted from biopanning results of phage-displayed random peptide libraries reported in 571 papers. The MimoDB database has been updated to the current version 2.0 very recently. In version 2.0, it has 15633 peptides collected from 849 papers and grouped into 1818 sets. Besides the core data on panning experiments and their results, broad background information on target, template, library and structure is included. An accompanied benchmark has also been compiled for bioinformaticians to develop and evaluate their new models, algorithms and programmes. In addition, the MimoDB database provides tools for simple and advanced searches, structure

visualization, BLAST and alignment view on the fly. The experimental biologists can easily use the database as a virtual control to exclude possible target-unrelated peptides. The MimoDB database is freely available.

ANTIGEN PRESENTATION

Antigen presentation describes a vital immune process which is essential for T cell immune response triggering. Because T cells recognise only fragmented antigens displayed on cell surfaces, antigen processing must occur before the antigen fragment, now bound to the major histocompatibility complex (MHC), is transported to the surface of the cell, a process known as presentation, where it can be recognized by a T cell receptor. If there has been an infection with viruses or bacteria, the cell will present an endogenous or exogenous peptide fragment derived from the antigen bound to MHC molecules. There are two types of MHC molecules which differ in the source of the antigens: MHC class I molecules (MHC-I) bind peptides from the cell cytosol, while peptides generated in the endocytic vesicles after internalisation are bound to MHC class II (MHC-II). Cellular membranes separate these two cellular environments - intracellular and extracellular. Each T cell can finally recognise only ten to hundreds copies of a unique sequence of a single peptide among thousands of other peptides presented on the very same cell because MHC molecule in one cell can bind quite a large range of peptides.

Presentation of Intracellular Antigens: Class I

Cytotoxic T cells (also known as T_c , killer T cell, or cytotoxic T-lymphocyte (CTL)) express CD8 coreceptor are a population of T cells that are specialised for inducing programmed cell death of other cells. Cytotoxic T cells regularly patrol all body cells to maintain the organismal homeostasis. Whenever they encounter signs of disease, caused for example by the presence of viruses or intracellular bacteria or a transformed tumor cell, they initiate processes to destroy the potentially harmful cell. All nucleated cells in the body (along with platelets) display class I major histocompatibility complex (MHC-I molecules). Antigens generated endogenously within these cells are bound to MHC-I molecules and presented on the cell surface. This antigen presentation pathway enables the immune system to detect transformed or infected cells displaying peptides from modified-self (mutated) or foreign proteins. In the presentation process, these proteins are mainly degraded into small peptides by cytosolic proteases in the proteasome, but there are also other cytoplasmic proteolytic pathways. Then, peptides are distributed to the endoplasmic reticulum (ER) via the action of heat shock proteins and the transporter associated with antigen processing (TAP) which translocates the cytosolic peptides into the ER lumen in an ATP-dependent transport mechanism. There are several ER chaperones involved in MHC-I assembly, such as calnexin, claretulin and tapasin. Peptides are loaded to MHC-I peptide binding groove between two alpha helices at the bottom of the $\alpha 1$ and $\alpha 2$ domains of the MHC class I molecule. After releasing from tapasin, peptide-MHC-I complexes (pMHC-I) exit the ER and are transported to the cell surface by exocytic vesicles. Naïve anti-viral T cells (CD8+) cannot directly eliminate transformed or infected cells. They have to be activated by the pMHC-I complexes of antigen-presenting cells (APCs). Here, antigen can be presented directly (as described above) or indirectly (cross-presentation) from virus-infected and non-infected cells. After the interaction between pMHC-I and TCR, in presence of costimulatory signals and/or cytokines, T cells are activated, migrate to the peripheral tissues and kill the target cells (infected or damaged cells) by inducing cytotoxicity. Cross-presentation is a special case in which MHC-I molecules are able to present extracellular antigens, usually displayed only by MHC-II molecules. This ability appears in several APCs, mainly plasmacytoid dendritic cells in tissues that stimulate CD8+ T cells directly. This process is essential when APCs are not directly infected, triggering local antiviral and anti-tumor immune responses immediately without trafficking the APCs in the local lymph nodes.

Presentation of Extracellular Antigens: Class II

Antigens from the extracellular space and sometimes also endogenous ones, are enclosed into endocytic vesicles and presented on the cell surface by MHC-II molecules to the helper T cells expressing CD4 molecule. Only APCs do express the class II of MHC molecules on the surface in large quantity, such as dendritic cells, B cells or macrophages, so expression of MHC-II molecules is more cell-specific than MHC-I. APCs usually internalise exogenous antigens by endocytosis, but also by pinocytosis, macroautophagy, endosomal microautophagy or chaperone-mediated autophagy. In the first case, after internalisation, the antigens are enclosed in vesicles called endosomes. There are three compartments involved in this antigen presentation pathway: early endosomes, late endosomes or endolysosomes and lysosomes, where antigens are hydrolyzed by lysosome-associated enzymes (acid-dependent hydrolases, glycosidases, proteases, lipases). This process is favored by gradual reduction of the pH. The main proteases in endosomes are cathepsins and the result is the degradation of the antigens into oligopeptides. MHC-II molecules are transported from the ER to the MHC class II loading compartment together with the protein Invariant chain (Ii, CD74). A non-classical MHC-II molecule (HLA-DO and HLA-DM) catalyses the exchange of part of the CD74 (CLIP peptide) with the peptide antigen. Peptide-MHC-II complexes (pMHC-II) are transported to the plasma membrane and the processed antigen is presented to the helper T cells in the lymph nodes. APCs undergo a process of maturation while migrating, via chemotactic signals, to lymphoid tissues, in which they lose the phagocytic capacity and develop an increased ability to communicate with T-cells by antigen-presentation. As well as in CD8+ cytotoxic T cells, APCs need pMHC-II and additional costimulatory signals to fully activate naive T helper cells. Alternative pathway of endogenous antigen processing and presentation over MHC-II molecules exists in medullary thymic epithelial cells (mTEC) via the process of autophagy. It is important for the process of central tolerance of T cells in particular the negative selection of autoreactive clones. Random gene expression of the whole genome is achieved via the action of AIRE and a self-digestion of the expressed molecules presented on both MHC-I and MHC-II molecules.

Presentation of Native Intact Antigens to B cells

B-cell receptors on the surface of B cells bind to intact native and undigested antigens of structural nature, rather than to a linear sequence of a peptide which has been digested into small fragments and presented by MHC molecules. Large complexes of intact antigen are presented in lymph nodes to B cells by follicular dendritic cells in the form of immune complexes. Some APCs expressing comparatively lower levels of lysosomal enzymes are thus less likely to digest the antigen they have captured before presenting it to B cells.

Dendritic Cell

Dendritic cells (DCs) are antigen-presenting cells (also known as *accessory cells*) of the mammalian immune system. Their main function is to process antigen material and present it on the cell surface to the T cells of the immune system. They act as messengers between the innate and the adaptive immune systems. Dendritic cells are present in those tissues that are in contact with the external environment, such as the skin (where there is a specialized dendritic cell type called the Langerhans cell) and the inner lining of the nose, lungs, stomach and intestines.

They can also be found in an immature state in the blood. Once activated, they migrate to the lymph nodes where they interact with T cells and B cells to initiate and shape the adaptive immune response. At certain development stages they grow branched projections, the *dendrites* that give the cell its name (δενδρον or déndron being Greek for “tree”). While similar in appearance, these are structures distinct from the dendrites of neurons. Immature dendritic cells are also called veiled cells, as they possess large cytoplasmic ‘veils’ rather than dendrites.

History

Dendritic cells were first described by Paul Langerhans (hence “Langerhans cells”) in the late nineteenth century. The term “dendritic cells” was coined in 1973 by Ralph M. Steinman and Zanvil A. Cohn. For discovering the central role of dendritic cells in the adaptive immune response, Steinman was awarded the Albert Lasker Award for Basic Medical Research in 2007 and the Nobel Prize in Physiology or Medicine in 2011.

Types

The morphology of dendritic cells results in a very large surface-to-volume ratio. That is, the dendritic cell has a very large surface area compared to the overall cell volume.

In vivo – primate

The most common division of dendritic cells is “myeloid” vs. “plasmacytoid” (lymphoid):

Name	Description	Secretion	Toll-like receptors
Conventional dendritic cell (previously called Myeloid dendritic cell) (cDC or mDC)	Most similar to monocytes. mDC are made up of at least two subsets: (1) The more common mDC-1, which is a major stimulator of T cells (2) the extremely rare mDC-2, which may have a function in fighting wound infection	Interleukin 12 (IL-12)	TLR 2, TLR 4
Plasmacytoid dendritic cell (pDC)	Look like plasma cells, but have certain characteristics similar to myeloid dendritic cells.	Can produce high amounts of interferon-alpha and were previously called interferon-producing cells.	TLR 7, TLR 9

The markers BDCA-2, BDCA-3, and BDCA-4 can be used to discriminate among the types.

Lymphoid and myeloid DCs evolve from lymphoid and myeloid precursors, respectively, and thus are of hematopoietic origin. By contrast, follicular dendritic cells (FDC) are probably of mesenchymal rather than hematopoietic origin and do not express MHC class II, but are so named because they are located in lymphoid follicles and have long “dendritic” processes.

In Blood

The blood DCs are typically identified and enumerated in flow cytometry. Three types of DCs have been defined in human blood: the CD1c+ myeloid DCs, the CD141+ myeloid DCs and the CD303+ plasmacytoid DCs. This represents the nomenclature proposed by the nomenclature committee of the International Union of Immunological Societies.

Dendritic cells that circulate in blood do not have all the typical features of their counterparts in tissue, *i.e.* they are less mature and have no dendrites. Still, they can perform complex functions including chemokine-production (in CD1c+ myeloid DCs), cross-presentation (in CD141+ myeloid DCs), and IFNalpha production (in CD303+ plasmacytoid DCs).

In Vitro

In some respects, dendritic cells cultured *in vitro* do not show the same behaviour or capability as dendritic cells isolated *ex vivo*. Nonetheless, they are often used for research as they are still much more readily available than genuine DCs.

- Mo-DC or MDDC refers to cells matured from monocytes
- HP-DC refers to cells derived from hematopoietic progenitor cells.

Nonprimate

While humans and non-human primates such as rhesus macaques appear to have DCs divided into these groups, other species (such as the mouse) have different subdivisions of DCs.

Life Cycle

Formation of Immature Cells and their Maturation

Dendritic cells are derived from hematopoietic bone marrow progenitor cells. These progenitor cells initially transform into immature dendritic cells. These cells are characterized by high endocytic activity and low T-cell activation potential. Immature dendritic cells constantly sample the surrounding environment for pathogens such as viruses and bacteria. This is done through pattern recognition receptors (PRRs) such as the toll-like receptors (TLRs). TLRs recognize specific chemical signatures found on subsets of pathogens. Immature dendritic cells may also phagocytose small quantities of membrane from live own cells, in a process called nibbling. Once they have come into contact with a presentable antigen, they become activated into mature dendritic cells and begin to migrate to the lymph node. Immature dendritic cells phagocytose pathogens and degrade their proteins into small pieces and upon maturation present those fragments at their cell surface using MHC molecules. Simultaneously, they upregulate cell-surface receptors that act as co-receptors in T-cell activation such as CD80 (B7.1), CD86 (B7.2), and CD40 greatly enhancing their ability to activate T-cells. They also upregulate CCR7, a chemotactic receptor that induces the dendritic cell to travel through the blood stream to the spleen or through the lymphatic system to a lymph node. Here they act as antigen-presenting cells: they activate helper T-cells and killer T-cells as well as B-cells by presenting them with antigens derived from the pathogen, alongside non-antigen specific costimulatory signals. Dendritic cells can also induce T-cell tolerance (unresponsiveness). Certain C-type lectin receptors (CLRs) on the surface of dendritic cells, some functioning as PRRs, help instruct dendritic cells as to when it is appropriate to induce immune tolerance rather than lymphocyte activation.

Every helper T-cell is specific to one particular antigen. Only professional antigen-presenting cells (macrophages, B lymphocytes, and dendritic cells) are able to activate a resting helper T-cell when the matching antigen is presented. However, in non-lymphoid organs, macrophages and B cells can only activate memory T cells whereas dendritic cells can activate both memory and naive T cells, and are the most potent of all the antigen-presenting cells. In the lymph node and secondary lymphoid organs, all three cell types can activate naive T cells. Whereas mature dendritic cells are able to activate antigen-specific naive CD8 T cells, the formation of CD8 memory T cells requires the interaction of dendritic cells with CD4 helper T cells. This help from CD4 T cells additionally activates the matured dendritic cells and licenses them to efficiently induce CD8 memory T cells, which are also able to be expanded a second time. For this activation of dendritic cells, concurrent interaction of all three cell types, namely CD4 T helper cells, CD8 T cells and dendritic cells, seems to be required.

As mentioned above, mDC probably arise from monocytes, white blood cells which circulate in the body and, depending on the right signal, can turn into either dendritic cells or macrophages. The monocytes in turn are formed from stem cells in the bone marrow. Monocyte-derived dendritic cells can be generated in vitro from peripheral blood mononuclear cell (PBMCs). Plating of PBMCs in a tissue culture flask permits adherence of monocytes. Treatment of these monocytes with interleukin 4 (IL-4) and granulocyte-macrophage colony stimulating factor (GM-CSF) leads to differentiation to immature dendritic cells (iDCs) in about a week. Subsequent treatment with tumor necrosis factor (TNF) further differentiates the iDCs into mature dendritic cells. Monocytes can be induced to differentiate into dendritic cells by a self-peptide Ep1.B derived from apolipoprotein E. These are primarily tolerogenic plasmacytoid dendritic cells.

Life Span

Activated macrophages have a lifespan of only a few days though new evidence suggests that it could be extended to weeks rather than days. The lifespan of activated dendritic cells, while somewhat varying according to type and origin, is of a similar order of magnitude, but immature dendritic cells seem to be able to exist in an inactivated state for much longer.

Research Challenges

The exact genesis and development of the different types and subsets of dendritic cells and their interrelationship is only marginally understood at the moment, as dendritic cells are so rare and difficult to isolate that only in recent years they have become subject of focused research. Distinct surface antigens that characterize dendritic cells have only become known from 2000 on; before that, researchers had to work with a 'cocktail' of several antigens which, used in combination, result in isolation of cells with characteristics unique to DCs.

Cytokines

The dendritic cells are constantly in communication with other cells in the body. This communication can take the form of direct cell–cell contact based on the interaction of cell-surface proteins. An example of this includes the interaction of the membrane proteins of the B7 family of the dendritic cell with CD28 present on the lymphocyte. However, the cell–cell interaction can also take place at a distance via cytokines. For example, stimulating dendritic cells *in vivo* with microbial extracts causes the dendritic cells to rapidly begin producing IL-12. IL-12 is a signal that helps send naive CD4 T cells towards a Th1 phenotype. The ultimate consequence is priming and activation of the immune system for attack against the antigens which the dendritic cell presents on its surface. However, there are differences in the cytokines produced depending on the type of dendritic cell. The plasmacytoid DC has the ability to produce huge amounts of type-1 IFNs, which recruit more activated macrophages to allow phagocytosis.

Disease

HIV Infection

HIV, which causes AIDS, can bind to dendritic cells via various receptors expressed on the cell. The best studied example is DC-SIGN (usually on MDC subset 1, but also on other subsets under certain conditions; since not all dendritic cell subsets express DC-SIGN, its exact role in sexual HIV-1 transmission is not clear). When the dendritic cell takes up HIV and then travels to the lymph node, the virus can be transferred to helper CD4+ T-cells, contributing to the developing infection. This infection of dendritic cells by HIV explains one mechanism by which the virus could persist after prolonged HAART. Many other viruses, such as the SARS virus seems to use DC-SIGN to 'hitchhike' to its target cells. However, most work with virus binding to DC-SIGN expressing cells has been conducted using *in vitro* derived cells such as moDCs. The physiological role of DC-SIGN *in vivo* is more difficult to ascertain.

Autoimmunity

Altered function of dendritic cells is also known to play a major or even key role in allergy and autoimmune diseases like lupus erythematosus and inflammatory bowel diseases (Crohn's disease and ulcerative colitis).

In Animals other than Humans

The above applies to humans. In other organisms, the function of dendritic cells can differ slightly. For example, in brown rats (but not mice), a subset of dendritic cells exists that displays pronounced killer cell-like

activity, apparently through its entire lifespan. However, the principal function of dendritic cells as known to date is always to act as an immune sentinel. They survey the body and collect information relevant to the immune system, they are then able to instruct and direct the adaptive arms to respond to challenges. In addition, an immediate precursor to myeloid and lymphoid dendritic cells of the spleen has been identified. This precursor, termed pre-DC, lacks MHC class II surface expression, and is distinct from monocytes, which primarily give rise to DCs in non-lymphoid tissues. Dendritic cells have also been found in chickens and turtles.

Macrophage

Macrophages are a type of white blood cell, of the immune system, that engulfs and digests cellular debris, foreign substances, microbes, cancer cells, and anything else that does not have the type of proteins specific to healthy body cells on its surface in a process called phagocytosis. These large phagocytes are found in essentially all tissues, where they patrol for potential pathogens by amoeboid movement. They take various forms (with various names) throughout the body (*e.g.*, histiocytes, Kupffer cells, alveolar macrophages, microglia, and others), but all are part of the mononuclear phagocyte system. Besides phagocytosis, they play a critical role in nonspecific defence (innate immunity) and also help initiate specific defence mechanisms (adaptive immunity) by recruiting other immune cells such as lymphocytes. For example, they are important as antigen presenters to T cells. In humans, dysfunctional macrophages cause severe diseases such as chronic granulomatous disease that result in frequent infections. Beyond increasing inflammation and stimulating the immune system, macrophages also play an important anti-inflammatory role and can decrease immune reactions through the release of cytokines. Macrophages that encourage inflammation are called M1 macrophages, whereas those that decrease inflammation and encourage tissue repair are called M2 macrophages. This difference is reflected in their metabolism; M1 macrophages have the unique ability to metabolize arginine to the “killer” molecule nitric oxide, whereas rodent M2 macrophages have the unique ability to metabolize arginine to the “repair” molecule ornithine. However, this dichotomy has been recently questioned as further complexity has been discovered. Human macrophages are about 21 micrometres (0.00083 in) in diameter and are produced by the differentiation of monocytes in tissues. They can be identified using flow cytometry or immunohistochemical staining by their specific expression of proteins such as CD14, CD40, CD11b, CD64, F4/80 (mice)/EMR1 (human), lysozyme M, MAC-1/MAC-3 and CD68. Macrophages were first discovered by Élie Metchnikoff, a Russian zoologist, in 1884.

Structure

Types

A majority of macrophages are stationed at strategic points where microbial invasion or accumulation of foreign particles is likely to occur. These cells together as a group are known as the mononuclear phagocyte system and were previously known as the reticuloendothelial system. Each type of macrophage, determined by its location, has a specific name:

Cell Name	Anatomical Location
Adipose tissue macrophages	Adipose tissue
Monocytes	Bone marrow/blood
Kupffer cells	Liver
Sinus histiocytes	Lymph nodes
Alveolar macrophages (dust cells)	Pulmonary alveoli of lungs
Tissue macrophages (histiocytes) leading to giant cells	Connective tissue
Microglia	Central nervous system

Hofbauer cells	Placenta
Intraglomerular mesangial cells	Kidney
Osteoclasts	Bone
Epithelioid cells	Granulomas
Red pulp macrophages (Sinusoidal lining cells)	Red pulp of spleen
Peritoneal macrophages	Peritoneal cavity
LysoMac	Peyer's patch

Investigations concerning Kupffer cells are hampered because in humans, Kupffer cells are only accessible for immunohistochemical analysis from biopsies or autopsies. From rats and mice, they are difficult to isolate, and after purification, only approximately 5 million cells can be obtained from one mouse. Macrophages can express paracrine functions within organs that are specific to the function of that organ. In the testis for example, macrophages have been shown to be able to interact with Leydig cells by secreting 25-hydroxycholesterol, an oxysterol that can be converted to testosterone by neighbouring Leydig cells. Also, testicular macrophages may participate in creating an immune privileged environment in the testis, and in mediating infertility during inflammation of the testis.

Cardiac resident macrophages participate in electrical conduction via gap junction communication with cardiac myocytes. Macrophages can be classified on basis of the fundamental function and activation. According to this grouping there are classically activated macrophages, wound-healing macrophages (alternatively activated macrophages) and regulatory macrophages (Mregs).

Development

Macrophages that reside in adult healthy tissues either derive from circulating monocytes or are established before birth and then maintained during adult life independently of monocytes. By contrast, most of the macrophages that accumulate at diseased sites typically derive from circulating monocytes. When a monocyte enters damaged tissue through the endothelium of a blood vessel, a process known as leukocyte extravasation, it undergoes a series of changes to become a macrophage. Monocytes are attracted to a damaged site by chemical substances through chemotaxis, triggered by a range of stimuli including damaged cells, pathogens and cytokines released by macrophages already at the site. At some sites such as the testis, macrophages have been shown to populate the organ through proliferation. Unlike short-lived neutrophils, macrophages survive longer in the body, up to several months.

Function

Phagocytosis

Macrophages are professional phagocytes and are highly specialized in removal of dying or dead cells and cellular debris. This role is important in chronic inflammation, as the early stages of inflammation are dominated by neutrophils, which are ingested by macrophages if they come of age. The neutrophils are at first attracted to a site, where they proliferate, before they are phagocytized by the macrophages. When at the site, the first wave of neutrophils, after the process of aging and after the first 48 hours, stimulate the appearance of the macrophages whereby these macrophages will then ingest the aged neutrophils. The removal of dying cells is, to a greater extent, handled by *fixed macrophages*, which will stay at strategic locations such as the lungs, liver, neural tissue, bone, spleen and connective tissue, ingesting foreign materials such as pathogens and recruiting additional macrophages if needed. When a macrophage ingests a pathogen, the pathogen becomes trapped in a phagosome, which then fuses with a lysosome. Within the phagolysosome, enzymes and toxic peroxides digest the pathogen. However, some bacteria, such as *Mycobacterium tuberculosis*, have become resistant to these methods of digestion. Typhoidal *Salmonellae* induce their own phagocytosis by host macrophages in vivo, and inhibit

digestion by lysosomal action, thereby using macrophages for their own replication and causing macrophage apoptosis. Macrophages can digest more than 100 bacteria before they finally die due to their own digestive compounds.

Role in Adaptive Immunity

Macrophages are versatile cells that play many roles. As scavengers, they rid the body of worn-out cells and other debris. Along with dendritic cells, they are foremost among the cells that present antigens, a crucial role in initiating an immune response. As secretory cells, monocytes and macrophages are vital to the regulation of immune responses and the development of inflammation; they produce a wide array of powerful chemical substances (monokines) including enzymes, complement proteins, and regulatory factors such as interleukin-1. At the same time, they carry receptors for lymphokines that allow them to be “activated” into single-minded pursuit of microbes and tumour cells. After digesting a pathogen, a macrophage will present the antigen (a molecule, most often a protein found on the surface of the pathogen and used by the immune system for identification) of the pathogen to the corresponding helper T cell. The presentation is done by integrating it into the cell membrane and displaying it attached to an MHC class II molecule (MHCII), indicating to other white blood cells that the macrophage is not a pathogen, despite having antigens on its surface. Eventually, the antigen presentation results in the production of antibodies that attach to the antigens of pathogens, making them easier for macrophages to adhere to with their cell membrane and phagocytose. In some cases, pathogens are very resistant to adhesion by the macrophages. The antigen presentation on the surface of infected macrophages (in the context of MHC class II) in a lymph node stimulates TH1 (type 1 helper T cells) to proliferate (mainly due to IL-12 secretion from the macrophage). When a B-cell in the lymph node recognizes the same unprocessed surface antigen on the bacterium with its surface bound antibody, the antigen is endocytosed and processed. The processed antigen is then presented in MHCII on the surface of the B-cell. T cells that express the T cell receptor which recognizes the antigen-MHCII complex (with co-stimulatory factors- CD40 and CD40L) cause the B-cell to produce antibodies that help opsonisation of the antigen so that the bacteria can be better cleared by phagocytes.

Macrophages provide yet another line of defence against tumor cells and somatic cells infected with fungus or parasites. Once a T cell has recognized its particular antigen on the surface of an aberrant cell, the T cell becomes an activated effector cell, producing chemical mediators known as lymphokines that stimulate macrophages into a more aggressive form.

Macrophage subtypes:

- There are several activated forms of macrophages. In spite of a spectrum of ways to activate macrophages, there are two main groups designated M1 and M2. M1 macrophages: as mentioned earlier (previously referred to as classically activated macrophages), M1 “killer” macrophages are activated by LPS and IFN-gamma, and secrete high levels of IL-12 and low levels of IL-10. M1 macrophages have pro-inflammatory, bactericidal, and phagocytic functions. In contrast, the M2 “repair” designation (also referred to as alternatively activated macrophages) broadly refers to macrophages that function in constructive processes like wound healing and tissue repair, and those that turn off damaging immune system activation by producing anti-inflammatory cytokines like IL-10. M2 is the phenotype of resident tissue macrophages, and can be further elevated by IL-4. M2 macrophages produce high levels of IL-10, TGF-beta and low levels of IL-12. Tumor-associated macrophages are mainly of the M2 phenotype, and seem to actively promote tumor growth.
- Macrophages exist in a variety of phenotypes which are determined by the role they play in wound maturation. Phenotypes can be predominantly separated into two major categories; M1 and M2. M1 macrophages are activated by four key mediators: interferon- γ (IFN- γ), tumor necrosis factor (TNF),

and damage associated pattern molecules (DAMPs). These mediator molecules create a pro-inflammatory response that in return produce pro-inflammatory cytokines like Interleukin-6 and TNF. These cytokines are essential in the initial process of wound healing. Unlike M1 macrophages, M2's secrete an anti-inflammatory response via the addition of Interleukin-4 or Interleukin-13. M2 cells are divided into four major types based on their roles: M2a, M2b, M2c, and M2d. How M2 phenotypes are determined is still up for discussion but studies have shown that their environment allows them to adjust to whichever phenotype is most appropriate to efficiently heal the wound.

- While M1 macrophages are the dominating phenotype observed in the early stages of inflammation, as the wound ages there is a significant decrease in M1 phenotype and an increase of M2 macrophages at the site. If this shift failed to occur, there would be prolonged inflammation. M2 cells are needed for production of collagen at the wound site. They are needed for revascularization and reepithelisation. It was previously thought that an increase of M2 macrophages may decrease the time it takes for wound closure. However, studies show that rate of wound closure is not affected by an increase in M2 cells.
- M2 macrophages are needed for vascular stability. They produce vascular epithelial growth factor-A and TGF- β 1. There is a phenotype shift from M1 to M2 macrophages in acute wounds, however this shift is impaired for chronic wounds. This dysregulation results in insufficient M2 macrophages and its corresponding growth factors that aid in wound repair. With a lack of these growth factors/anti-inflammatory cytokines and an overabundance of pro-inflammatory cytokines from M1 macrophages chronic wounds are unable to heal in a timely manner. Normally, after neutrophils eat debris/pathogens they perform apoptosis and are removed. At this point, inflammation is not needed and M1 undergoes a switch to M2 (anti-inflammatory). However, dysregulation occurs as the M1 macrophages are unable/do not phagocytose neutrophils that have undergone apoptosis leading to increased macrophage migration and inflammation.
- Both M1 and M2 macrophages play a role in promotion of atherosclerosis. M1 macrophages promote atherosclerosis by inflammation. M2 macrophages can remove cholesterol from blood vessels, but when the cholesterol is oxidized, the M2 macrophages become apoptotic foam cells contributing to the atheromatous plaque of atherosclerosis.

Role in Muscle Regeneration

The first step to understanding the importance of macrophages in muscle repair, growth, and regeneration is that there are two “waves” of macrophages with the onset of damageable muscle use – subpopulations that do and do not directly have an influence on repairing muscle. The initial wave is a phagocytic population that comes along during periods of increased muscle use that are sufficient to cause muscle membrane lysis and membrane inflammation, which can enter and degrade the contents of injured muscle fibers. These early-invading, phagocytic macrophages reach their highest concentration about 24 hours following the onset of some form of muscle cell injury or reloading.

Their concentration rapidly declines after 48 hours. The second group is the non-phagocytic types that are distributed near regenerative fibers. These peak between two and four days and remain elevated for several days during the hopeful muscle rebuilding. The first subpopulation has no direct benefit to repairing muscle, while the second non-phagocytic group does. It is thought that macrophages release soluble substances that influence the proliferation, differentiation, growth, repair, and regeneration of muscle, but at this time the factor that is produced to mediate these effects is unknown. It is known that macrophages' involvement in promoting tissue repair is not muscle specific; they accumulate in numerous tissues during the healing process phase following injury.

Role in Wound Healing

Macrophages are essential for wound healing. They replace polymorphonuclear neutrophils as the predominant cells in the wound by day two after injury. Attracted to the wound site by growth factors released by platelets and other cells, monocytes from the bloodstream enter the area through blood vessel walls. Numbers of monocytes in the wound peak one to one and a half days after the injury occurs. Once they are in the wound site, monocytes mature into macrophages. The spleen contains half the body's monocytes in reserve ready to be deployed to injured tissue. The macrophage's main role is to phagocytize bacteria and damaged tissue, and they also debride damaged tissue by releasing proteases. Macrophages also secrete a number of factors such as growth factors and other cytokines, especially during the third and fourth post-wound days. These factors attract cells involved in the proliferation stage of healing to the area. Macrophages may also restrain the contraction phase. Macrophages are stimulated by the low oxygen content of their surroundings to produce factors that induce and speed angiogenesis and they also stimulate cells that re-epithelialize the wound, create granulation tissue, and lay down a new extracellular matrix. By secreting these factors, macrophages contribute to pushing the wound healing process into the next phase.

Role in Limb Regeneration

Scientists have elucidated that as well as eating up material debris, macrophages are involved in the typical limb regeneration in the salamander. They found that removing the macrophages from a salamander resulted in failure of limb regeneration and a scarring response.

Role in Iron Homeostasis

As described above, macrophages play a key role in removing dying or dead cells and cellular debris. Erythrocytes have a lifespan on average of 120 days and so are constantly being destroyed by macrophages in the spleen and liver. Macrophages will also engulf macromolecules, and so play a key role in the pharmacokinetics of parenteral irons. The iron that is released from the haemoglobin is either stored internally in ferritin or is released into the circulation via ferroportin. In cases where systemic iron levels are raised, or where inflammation is present, raised levels of hepcidin act on macrophage ferroportin channels, leading to iron remaining within the macrophages.

Clinical Significance

Due to their role in phagocytosis, macrophages are involved in many diseases of the immune system. For example, they participate in the formation of granulomas, inflammatory lesions that may be caused by a large number of diseases. Some disorders, mostly rare, of ineffective phagocytosis and macrophage function have been described, for example.

As a Host for Intracellular Pathogens

In their role as a phagocytic immune cell macrophages are responsible for engulfing pathogens to destroy them. Some pathogens subvert this process and instead live inside the macrophage. This provides an environment in which the pathogen is hidden from the immune system and allows it to replicate. Diseases with this type of behaviour include tuberculosis (caused by *Mycobacterium tuberculosis*) and leishmaniasis (caused by *Leishmania* species). In order to minimize the possibility of becoming the host of an intracellular bacteria, macrophages have evolved defence mechanisms such as induction of nitric oxide and reactive oxygen intermediates, which are toxic to microbes. Macrophages have also evolved the ability to restrict the microbe's nutrient supply and induce autophagy.

Tuberculosis:

- Once engulfed by a macrophage, the causative agent of tuberculosis, *Mycobacterium tuberculosis*, avoids cellular defenses and uses the cell to replicate.

Leishmaniasis:

- Upon phagocytosis by a macrophage, the *Leishmania* parasite finds itself in a phagocytic vacuole. Under normal circumstances, this phagocytic vacuole would develop into a lysosome and its contents would be digested. *Leishmania* alter this process and avoid being destroyed; instead, they make a home inside the vacuole.

Chikungunya:

- Infection of macrophages in joints is associated with local inflammation during and after the acute phase of *Chikungunya* (caused by CHIKV or Chikungunya virus).

Others:

- Adenovirus (most common cause of pink eye) can remain latent in a host macrophage, with continued viral shedding 6–18 months after initial infection.
- *Brucella spp.* can remain latent in a macrophage via inhibition of phagosome–lysosome fusion; causes brucellosis (undulant fever).
- *Legionella pneumophila*, the causative agent of Legionnaires' disease, also establishes residence within macrophages.

Heart Disease

Macrophages are the predominant cells involved in creating the progressive plaque lesions of atherosclerosis. Focal recruitment of macrophages occurs after the onset of acute myocardial infarction. These macrophages function to remove debris, apoptotic cells and to prepare for tissue regeneration.

HIV Infection

Macrophages also play a role in human immunodeficiency virus (HIV) infection. Like T cells, macrophages can be infected with HIV, and even become a reservoir of ongoing virus replication throughout the body. HIV can enter the macrophage through binding of gp120 to CD4 and second membrane receptor, CCR5 (a chemokine receptor). Both circulating monocytes and macrophages serve as a reservoir for the virus. Macrophages are better able to resist infection by HIV-1 than CD4+ T cells, although susceptibility to HIV infection differs among macrophage subtypes.

Cancer

Macrophages can contribute to tumor growth and progression by promoting tumor cell proliferation and invasion, fostering tumor angiogenesis and suppressing antitumor immune cells. Attracted to oxygen-starved (hypoxic) and necrotic tumor cells they promote chronic inflammation. Inflammatory compounds such as tumor necrosis factor (TNF)-alpha released by the macrophages activate the gene switch nuclear factor-kappa B. NF-κB then enters the nucleus of a tumor cell and turns on production of proteins that stop apoptosis and promote cell proliferation and inflammation. Moreover, macrophages serve as a source for many pro-angiogenic factors including vascular endothelial factor (VEGF), tumor necrosis factor-alpha (TNF-alpha), Macrophage colony-stimulating factor (M-CSF/CSF1) and IL-1 and IL-6 contributing further to the tumor growth. Macrophages have been shown to infiltrate a number of tumors. Their number correlates with poor prognosis in certain cancers including cancers of breast, cervix, bladder and brain. Tumor-associated macrophages (TAMs) are thought to acquire an M2 phenotype, contributing to tumor growth and progression. Some tumors can also produce factors, including M-CSF/CSF1, MCP-1/CCL2 and Angiotensin II, that trigger the amplification and mobilization of macrophages in tumors. Research in various study models suggests that macrophages can

sometimes acquire anti-tumor functions. For example, macrophages may have cytotoxic activity to kill tumor cells directly; also the co-operation of T-cells and macrophages is important to suppress tumors. This co-operation involves not only the direct contact of T-cell and macrophage, with antigen presentation, but also includes the secretion of adequate combinations of cytokines, which enhance T-cell antitumor activity. Recent study findings suggest that by forcing IFN- α expression in tumor-infiltrating macrophages, it is possible to blunt their innate protumoral activity and reprogram the tumor microenvironment towards more effective dendritic cell activation and immune effector cell cytotoxicity. Additionally, subcapsular sinus macrophages in tumor-draining lymph nodes can suppress cancer progression by containing the spread of tumor-derived materials.

Cancer Therapy

Experimental studies indicate that macrophages can affect all therapeutic modalities, including surgery, chemotherapy, radiotherapy, immunotherapy and targeted therapy. Macrophages can influence treatment outcomes both positively and negatively. Macrophages can be protective in different ways: they can remove dead tumor cells (in a process called phagocytosis) following treatments that kill these cells; they can serve as drug depots for some anticancer drugs; they can also be activated by some therapies to promote antitumor immunity. Macrophages can also be deleterious in several ways: for example they can suppress various chemotherapies, radiotherapies and immunotherapies. Because macrophages can regulate tumor progression, therapeutic strategies to reduce the number of these cells, or to manipulate their phenotypes, are currently being tested in cancer patients.

Obesity

Increased number of pro-inflammatory macrophages within obese adipose tissue contributes to obesity complications including insulin resistance and diabetes type 2. Within the fat (adipose) tissue of CCR2 deficient mice, there is an increased number of eosinophils, greater alternative Macrophage activation, and a propensity towards type 2 cytokine expression. Furthermore, this effect was exaggerated when the mice became obese from a high fat diet.

Intestinal Macrophages

Though very similar in structure to tissue macrophages, intestinal macrophages have evolved specific characteristics and functions given their natural environment, which is in the digestive tract. Macrophages and intestinal macrophages have high plasticity causing their phenotype to be altered by their environments. Like macrophages, intestinal macrophages are differentiated monocytes, though intestinal macrophages have to coexist with the microbiome in the intestines. This is a challenge considering the bacteria found in the gut are not recognized as “self” and could be potential targets for phagocytosis by the macrophage.

To prevent the destruction of the gut bacteria, intestinal macrophages have developed key differences compared to other macrophages. Primarily, intestinal macrophages do not induce inflammatory responses. Whereas tissue macrophages release various inflammatory cytokines, such as IL-1, IL-6 and TNF- α , intestinal macrophages do not produce or secrete inflammatory cytokines. This change is directly caused by the intestinal macrophages environment. Surrounding intestinal epithelial cells release TGF- β , which induces the change from proinflammatory macrophage to noninflammatory macrophage. Even though the inflammatory response is downregulated in intestinal macrophages, phagocytosis is still carried out. There is no drop off in phagocytosis efficiency as intestinal macrophages are able to effectively phagocytize the bacteria, *S. typhimurium* and *E. coli*, but intestinal macrophages still do not release cytokines, even after phagocytosis. Also, intestinal macrophages do not express lipopolysaccharide (LPS), IgA, or IgG receptors. The lack of LPS receptors is important for the gut as the intestinal macrophages do not detect the microbe-associated molecular patterns (MAMPS/PAMPS) of the intestinal microbiome. Nor do they express IL-2 and IL-3 growth factor receptors.

Intestinal Macrophages Role in Disease

Intestinal macrophages have been shown to play a role in inflammatory bowel disease (IBD), such as Crohn's Disease (CD) and Ulcerative Colitis (UC). In a healthy gut, intestinal macrophages limit the inflammatory response in the gut, but in a disease-state, intestinal macrophage numbers and diversity are altered. This leads to inflammation of the gut and disease symptoms of IBD. Intestinal macrophages are critical in maintaining gut homeostasis. The presence of inflammation or pathogen alters this homeostasis, and concurrently alters the intestinal macrophages. There has yet to be a determined mechanism for the alteration of the intestinal macrophages by recruitment of new monocytes or changes in the already present intestinal macrophages.

B cell

B cells, also known as B lymphocytes, are a type of white blood cell of the lymphocyte subtype. They function in the humoral immunity component of the adaptive immune system by secreting antibodies. Additionally, B cells present antigen (they are also classified as professional antigen-presenting cells (APCs)) and secrete cytokines. In mammals, B cells mature in the bone marrow, which is at the core of most bones. In birds, B cells mature in the bursa of Fabricius, a lymphoid organ. (The "B" from B cells comes from the name of this organ, where it was first discovered by Chang and Glick, and not from bone marrow as commonly believed.) B cells, unlike the other two classes of lymphocytes, T cells and natural killer cells, express B cell receptors (BCRs) on their cell membrane. BCRs allow the B cell to bind to a specific antigen, against which it will initiate an antibody response.

Development

B cells develop from hematopoietic stem cells (HSCs) that originate from bone marrow. HSCs first differentiate into multipotent progenitor (MPP) cells, then common lymphoid progenitor (CLP) cells. From here, their development into B cells occurs in several stages (shown in image to the right), each marked by various gene expression patterns and immunoglobulin H chain and L chain gene loci arrangements, the latter due to B cells undergoing V(D)J recombination as they develop. B cells undergo two types of selection while developing in the bone marrow to ensure proper development. Positive selection occurs through antigen-independent signaling involving both the pre-BCR and the BCR. If these receptors do not bind to their ligand, B cells do not receive the proper signals and cease to develop. Negative selection occurs through the binding of self-antigen with the BCR; If the BCR can bind strongly to self-antigen, then the B cell undergoes one of four fates: clonal deletion, receptor editing, anergy, or ignorance (B cell ignores signal and continues development). This negative selection process leads to a state of central tolerance, in which the mature B cells don't bind with self antigens present in the bone marrow. To complete development, immature B cells migrate from the bone marrow into the spleen as transitional B cells, passing through two transitional stages: T1 and T2. Throughout their migration to the spleen and after spleen entry, they are considered T1 B cells. Within the spleen, T1 B cells transition to T2 B cells. T2 B cells differentiate into either follicular (FO) B cells or marginal zone (MZ) B cells depending on signals received through the BCR and other receptors. Once differentiated, they are now considered mature B cells, or naive B cells. While immature and during the T1 phase, B cells express BCR of class IgG, but BCR expression changes to the classes IgM and IgD after transition into the T2 phase and while mature up to activation.

Activation

B cell activation occurs in the secondary lymphoid organs (SLOs), such as the spleen and lymph nodes. After B cells mature in the bone marrow, they migrate through the blood to SLOs, which receive a constant supply of antigen through circulating lymph. At the SLO, B cell activation begins when the B cell binds to an

antigen via its BCR. Of the three B cell subsets, FO B cells preferentially undergo T cell-dependent activation while MZ B cells and B1 B cells preferentially undergo T cell-independent activation. B cell activation is enhanced through the activity of CD21, a surface receptor in complex with surface proteins CD19 and CD81 (all three are collectively known as the B cell coreceptor complex). When a BCR binds an antigen tagged with a fragment of the C3 complement protein, CD21 binds the C3 fragment, co-ligates with the bound BCR, and signals are transduced through CD19 and CD81 to lower the activation threshold of the cell.

T Cell-dependent Activation

Antigens that activate B cells with the help of T-cell are known as T cell-dependent (TD) antigens and include foreign proteins. They are named as such because they are unable to induce a humoral response in organisms that lack T cells. B cell response to these antigens takes multiple days, though antibodies generated have a higher affinity and are more functionally versatile than those generated from T cell-independent activation. Once a BCR binds a TD antigen, the antigen is taken up into the B cell through receptor-mediated endocytosis, degraded, and presented to T cells as peptide pieces in complex with MHC-II molecules on the cell membrane. T helper (T_H) cells, typically follicular T helper (T_{FH}) cells, that were activated with the same antigen recognize and bind these MHC-II-peptide complexes through their T cell receptor (TCR). Following TCR-MHC-II-peptide binding, T cells express the surface protein CD40L as well as cytokines such as IL-4 and IL-21. CD40L serves as a necessary co-stimulatory factor for B cell activation by binding the B cell surface receptor CD40, which promotes B cell proliferation, immunoglobulin class switching, and somatic hypermutation as well as sustains T cell growth and differentiation. T cell-derived cytokines bound by B cell cytokine receptors also promote B cell proliferation, immunoglobulin class switching, and somatic hypermutation as well as guide differentiation. After B cells receive these signals, they are considered activated.

Now activated, B cells participate in a two-step differentiation process that yields both short-lived plasmablasts for immediate protection and long-lived plasma cells and memory B cells for persistent protection. The first step, known as the extrafollicular response, occurs outside lymphoid follicles but still in the SLO. During this step activated B cells proliferate, may undergo immunoglobulin class switching, and differentiate into plasmablasts that produce early, weak antibodies mostly of class IgM. The second step consists of activated B cells entering a lymphoid follicle and forming a germinal center (GC), which is a specialized microenvironment where B cells undergo extensive proliferation, immunoglobulin class switching, and affinity maturation directed by somatic hypermutation. These processes are facilitated by T_{FH} cells within the GC and generate both high-affinity memory B cells and long-lived plasma cells. Resultant plasma cells secrete large amounts of antibody and either stay within the SLO or, more preferentially, migrate to bone marrow.

T Cell-independent Activation

Antigens that activate B cells without T cell help are known as T cell-independent (TI) antigens and include foreign polysaccharides and unmethylated CpG DNA. They are named as such because they are able to induce a humoral response in organisms that lack T cells. B cell response to these antigens is rapid, though antibodies generated tend to have lower affinity and are less functionally versatile than those generated from T cell-dependent activation.

As with TD antigens, B cells activated by TI antigens need additional signals to complete activation, but instead of receiving them from T cells, they are provided either by recognition and binding of a common microbial constituent to toll-like receptors (TLRs) or by extensive crosslinking of BCRs to repeated epitopes on a bacterial cell. B cells activated by TI antigens go on to proliferate outside lymphoid follicles but still in SLOs (GCs do not form), possibly undergo immunoglobulin class switching, and differentiate into short-lived plasmablasts that produce early, weak antibodies mostly of class IgM, but also some populations of long-lived plasma cells.

Memory B Cell Activation

Memory B cell activation begins with the detection and binding of their target antigen, which is shared by their parent B cell. Some memory B cells can be activated without T cell help, such as certain virus-specific memory B cells, but others need T cell help. Upon antigen binding, the memory B cell takes up the antigen through receptor-mediated endocytosis, degrades it, and presents it to T cells as peptide pieces in complex with MHC-II molecules on the cell membrane. Memory T helper (T_H) cells, typically memory follicular T helper (T_{FH}) cells, that were derived from T cells activated with the same antigen recognize and bind these MHC-II-peptide complexes through their TCR. Following TCR-MHC-II-peptide binding and the relay of other signals from the memory T_{FH} cell, the memory B cell is activated and differentiates either into plasmablasts and plasma cells via an extrafollicular response or enter a germinal center reaction where they generate plasma cells and more memory B cells. It is unclear whether the memory B cells undergo further affinity maturation within these secondary GCs.

B Cell Types

- **Plasmablast** - A short-lived, proliferating antibody-secreting cell arising from B cell differentiation. Plasmablasts are generated early in an infection and their antibodies tend to have a weaker affinity towards their target antigen compared to plasma cell. Plasmablasts can result from T cell-independent activation of B cells or the extrafollicular response from T cell-dependent activation of B cells.
- **Plasma cell** - A long-lived, non-proliferating antibody-secreting cell arising from B cell differentiation. There is evidence that B cells first differentiate into a plasmablast-like cell, then differentiate into a plasma cell. Plasma cells are generated later in an infection and, compared to plasmablasts, have antibodies with a higher affinity towards their target antigen due to affinity maturation in the germinal center (GC) and produce more antibodies. Plasma cells typically result from the germinal center reaction from T cell-dependent activation of B cells, however they can also result from T cell-independent activation of B cells.
- **Lymphoplasmacytoid cell** - A cell with a mixture of B lymphocyte and plasma cell morphological features that is thought to be closely related to or a subtype of plasma cells. This cell type is found in pre-malignant and malignant plasma cell dyscrasias that are associated with the secretion of IgM monoclonal proteins; these dyscrasias include IgM monoclonal gammopathy of undetermined significance and Waldenström's macroglobulinemia.
- **Memory B cell** - Dormant B cell arising from B cell differentiation. Their function is to circulate through the body and initiate a stronger, more rapid antibody response (known as the anamnestic secondary antibody response) if they detect the antigen that had activated their parent B cell (memory B cells and their parent B cells share the same BCR, thus they detect the same antigen). Memory B cells can be generated from T cell-dependent activation through both the extrafollicular response and the germinal center reaction as well as from T cell-independent activation of B1 cells.
- **Follicular (FO) B Cell** (also known as a B-2 cell) - Most common type of B cell and, when not circulating through the blood, is found mainly in the lymphoid follicles of secondary lymphoid organs (SLOs). They are responsible for generating the majority of high-affinity antibodies during an infection.
- **Marginal zone (MZ) B cell** - Found mainly in the marginal zone of the spleen and serves as a first line of defence against blood-borne pathogens, as the marginal zone receives large amounts of blood from the general circulation. They can undergo both T cell-independent and T cell-dependent activation, but preferentially undergo T cell-independent activation.
- **B-1 cell** - Arises from a developmental pathway different from FO B cells and MZ B cells. In mice, they predominantly populate the peritoneal cavity and pleural cavity, generate natural antibodies

(antibodies produced without infection), defend against mucosal pathogens, and primarily exhibit T cell-independent activation. A true homologue of mouse B-1 cells has not been discovered in humans, though various cell populations similar to B-1 cells have been described.

- B-2 cell - FO B cells and MZ B cells.
- Regulatory B (Breg) cell - An immunosuppressive B cell type that stops the expansion of pathogenic, pro-inflammatory lymphocytes through the secretion of IL-10, IL-35, and TGF- β . Also, it promotes the generation of regulatory T (Treg) cells by directly interacting with T cells to skew their differentiation towards Tregs. No common Breg cell identity has been described and many Breg cell subsets sharing regulatory functions have been found in both mice and humans. It is currently unknown if Breg cell subsets are developmentally linked and how exactly differentiation into a Breg cell occurs. There is evidence showing that nearly all B cell types can differentiate into a Breg cell through mechanisms involving inflammatory signals and BCR recognition.

B Cell-related Pathology

Autoimmune disease can result from abnormal B cell recognition of self-antigens followed by the production of autoantibodies. Autoimmune diseases where disease activity is correlated with B cell activity include scleroderma, multiple sclerosis, systemic lupus erythematosus, type 1 diabetes, and rheumatoid arthritis. Malignant transformation of B cells and their precursors can cause a host of cancers, including chronic lymphocytic leukemia (CLL), acute lymphoblastic leukemia (ALL), hairy cell leukemia, follicular lymphoma, non-Hodgkin's lymphoma, Hodgkin's lymphoma, and plasma cell malignancies such as multiple myeloma, Waldenström's macroglobulinemia, and certain forms of amyloidosis.

Immunogen

An immunogen is an antigen or any substance that may be specifically bound by components of the immune system (antibody, lymphocytes). The term antigen arises from its ability to induce generation of antibodies (antigen = antibody generation). Despite the fact that all antigens are recognized by specific lymphocytes or by antibodies, not every antigen can evoke an immune response. Those antigens that are capable of inducing an immune response are said to be immunogenic and are called immunogens. An immunogen is any antigen that is capable of inducing humoral and/or cell-mediated immune response rather than immunological tolerance. This ability is called immunogenicity. Sometimes the term immunogen is used interchangeably with the term antigen. But only an immunogen can evoke an immune response. Generally, both are substances that are capable of generating antibodies (antigen) or stimulating immune responses (immunogen). We can define an immunogen as a complete antigen which is composed of the macromolecular carrier and epitopes (determinants) that can induce immune response.

An explicit example is a hapten. Haptens are low-molecular-weight compounds that may be bound by antibodies, but cannot elicit an immune response. Consequently, the haptens themselves are nonimmunogenic and they cannot evoke an immune response until they bind with a larger carrier immunogenic molecule. The hapten-carrier complex, unlike free hapten, can act as an immunogen and can induce an immune response. Until 1959, the terms immunogen and antigen were not distinguished.

Used Carrier Proteins

- *Keyhole Limpet Hemocyanin (KLH)*: It is copper-containing respiratory protein, isolated from keyhole limpets (*Megathura crenulata*). Because of its evolutionary distance from mammals, high molecular weight and complex structure it is usually immunogenic in vertebrate animals.

- *Concholepas Concholepas Hemocyanin (CCH)*: (also Blue Carrier Immunogenic Protein) It is alternative to KLH isolated from *Concholepas concholepas*. It has the similar immunogenic properties as KLH but better solubility and therefore better flexibility.
- *Bovine Serum Albumin (BSA)*: It is from the blood sera of cows and has similarly immunogenic properties as KLH or CCH. The cationized form of BSA (cBSA) is highly positively charged protein with significantly increased immunogenicity. This change possesses a greater number of possible conjugated antigens to the protein.
- *Ovalbumin (OVA)*: Also known as egg albumin, OVA is the main protein (60-75 per cent) found in hen egg white. OVA is soluble in Dimethyl Sulfoxide (DMSO), which enables the conjugation of haptens that are not soluble in aqueous buffers. The immune response can be enhanced using an adjuvant injected together with the immunogen.

Immunological Adjuvants

- An adjuvant (from Latin *adiuvare* – to help) is any substance, distinct from antigen, which enhances immune response by various mechanisms: recruiting of professional antigen-presenting cells (APCs) to the site of antigen exposure; increasing the delivery of antigens by delayed/slow release (depot generation); immunomodulation by cytokine production (selection of Th1 or Th2 response); inducing T-cell response (prolonged exposure of peptide-MHC complexes [signal 1] and stimulation of expression of T-cell-activating co-stimulators [signal 2] on the APCs' surface) and targeting (e. g. carbohydrate adjuvants which target lectin receptors on APCs). Adjuvants have been used as additives to improve vaccine efficiency since the 1920s. Generally, administration of adjuvants is used both in experimental immunology and in clinical settings to ensure a high quality/quantity memory-enhanced antibody response, where antigens must be prepared and delivered in a fashion that maximizes production of a specific immune response. Among commonly used adjuvants are Complete and Incomplete Freund's adjuvant and solutions of aluminum hydroxide or aluminum phosphate.

ANTIBODIES

An antibody (Ab), also known as an immunoglobulin (Ig), is a large, Y-shaped protein produced mainly by plasma cells that is used by the immune system to neutralize pathogens such as pathogenic bacteria and viruses. The antibody recognizes a unique molecule of the pathogen, called an antigen, via the Fab's variable region. Each tip of the "Y" of an antibody contains a paratope (analogous to a lock) that is specific for one particular epitope (similarly analogous to a key) on an antigen, allowing these two structures to bind together with precision. Using this binding mechanism, an antibody can *tag* a microbe or an infected cell for attack by other parts of the immune system, or can neutralize its target directly (for example, by inhibiting a part of a microbe that is essential for its invasion and survival). Depending on the antigen, the binding may impede the biological process causing the disease or may activate macrophages to destroy the foreign substance. The ability of an antibody to communicate with the other components of the immune system is mediated via its Fc region (located at the base of the "Y"), which contains a conserved glycosylation site involved in these interactions. The production of antibodies is the main function of the humoral immune system.

Antibodies are secreted by B cells of the adaptive immune system, mostly by differentiated B cells called plasma cells. Antibodies can occur in two physical forms, a soluble form that is secreted from the cell to be free in the blood plasma, and a membrane-bound form that is attached to the surface of a B cell and is referred to as the B-cell receptor (BCR). The BCR is found only on the surface of B cells and facilitates the activation of these cells and their subsequent differentiation into either antibody factories called plasma cells or memory B cells that will survive in the body and remember that same antigen so the B cells can respond

faster upon future exposure. In most cases, interaction of the B cell with a T helper cell is necessary to produce full activation of the B cell and, therefore, antibody generation following antigen binding. Soluble antibodies are released into the blood and tissue fluids, as well as many secretions to continue to survey for invading microorganisms. Antibodies are glycoproteins belonging to the immunoglobulin superfamily. They constitute most of the gamma globulin fraction of the blood proteins. They are typically made of basic structural units—each with two large heavy chains and two small light chains. There are several different types of antibody heavy chains that define the five different types of crystallisable fragments (Fc) that may be attached to the antigen-binding fragments. The five different types of Fc regions allow antibodies to be grouped into five *isotypes*. Each Fc region of a particular antibody isotype is able to bind to its specific Fc Receptor (except for IgD, which is essentially the BCR), thus allowing the antigen-antibody complex to mediate different roles depending on which FcR it binds. The ability of an antibody to bind to its corresponding FcR is further modulated by the structure of the glycan(s) present at conserved sites within its Fc region. The ability of antibodies to bind to FcRs helps to direct the appropriate immune response for each different type of foreign object they encounter. For example, IgE is responsible for an allergic response consisting of mast cell degranulation and histamine release. IgE's Fab paratope binds to allergic antigen, for example house dust mite particles, while its Fc region binds to Fc receptor ϵ . The allergen-IgE-FcR ϵ interaction mediates allergic signal transduction to induce conditions such as asthma.

Though the general structure of all antibodies is very similar, a small region at the tip of the protein is extremely variable, allowing millions of antibodies with slightly different tip structures, or antigen-binding sites, to exist. This region is known as the *hypervariable region*. Each of these variants can bind to a different antigen. This enormous diversity of antibody paratopes on the antigen-binding fragments allows the immune system to recognize an equally wide variety of antigens. The large and diverse population of antibody paratope is generated by random recombination events of a set of gene segments that encode different antigen-binding sites (or *paratopes*), followed by random mutations in this area of the antibody gene, which create further diversity. This recombinational process that produces clonal antibody paratope diversity is called V(D)J or VJ recombination. Basically, the antibody paratope is polygenic, made up of three genes, V, D, and J. Each paratope locus is also polymorphic, such that during antibody production, one allele of V, one of D, and one of J is chosen. These gene segments are then joined together using random genetic recombination to produce the paratope. The regions where the genes are randomly recombined together is the hyper variable region used to recognise different antigens on a clonal basis. Antibody genes also re-organize in a process called class switching that changes the one type of heavy chain Fc fragment to another, creating a different isotype of the antibody that retains the antigen-specific variable region. This allows a single antibody to be used by different types of Fc receptors, expressed on different parts of the immune system.

FORMS

The membrane-bound form of an antibody may be called a *surface immunoglobulin* (sIg) or a *membrane immunoglobulin* (mIg). It is part of the *B cell receptor* (BCR), which allows a B cell to detect when a specific antigen is present in the body and triggers B cell activation. The BCR is composed of surface-bound IgD or IgM antibodies and associated Ig- α and Ig- β heterodimers, which are capable of signal transduction. A typical human B cell will have 50,000 to 100,000 antibodies bound to its surface. Upon antigen binding, they cluster in large patches, which can exceed 1 micrometer in diameter, on lipid rafts that isolate the BCRs from most other cell signaling receptors.

These patches may improve the efficiency of the cellular immune response. In humans, the cell surface is bare around the B cell receptors for several hundred nanometers, which further isolates the BCRs from competing influences.

ANTIBODY–ANTIGEN INTERACTIONS

The antibody's paratope interacts with the antigen's epitope. An antigen usually contains different epitopes along its surface arranged discontinuously, and dominant epitopes on a given antigen are called determinants. Antibody and antigen interact by spatial complementarity (lock and key). The molecular forces involved in the Fab-epitope interaction are weak and non-specific – for example electrostatic forces, hydrogen bonds, hydrophobic interactions, and van der Waals forces. This means binding between antibody and antigen is reversible, and the antibody's affinity towards an antigen is relative rather than absolute. Relatively weak binding also means it is possible for an antibody to cross-react with different antigens of different relative affinities.

Often, once an antibody and antigen bind, they become an immune complex, which functions as a unitary object and can act as an antigen in its own right, being countered by other antibodies. Similarly, haptens are small molecules that provoke no immune response by themselves, but once they bind to proteins, the resulting complex or hapten-carrier adduct is antigenic.

ISOTYPES

Antibodies can come in different varieties known as isotypes or classes. In placental mammals there are five antibody isotypes known as IgA, IgD, IgE, IgG, and IgM. They are each named with an “Ig” prefix that stands for immunoglobulin, a name sometimes used interchangeably with antibody, and differ in their biological properties, functional locations and ability to deal with different antigens, as depicted in the table. The different suffixes of the antibody isotypes denote the different types of heavy chains the antibody contains, with each heavy chain class named alphabetically: α (alpha), γ (gamma), δ (delta), ϵ (epsilon), and μ (mu). This gives rise to IgA, IgG, IgD, IgE, and IgM, respectively. The antibody isotype of a B cell changes during cell development and activation. Immature B cells, which have never been exposed to an antigen, express only the IgM isotype in a cell surface bound form. The B lymphocyte, in this ready-to-respond form, is known as a “naive B lymphocyte.” The naive B lymphocyte expresses both surface IgM and IgD.

Class	Subclass	Description	Antibody complexes
IgA	2	Found in mucosal areas, such as the gut, respiratory tract and urogenital tract, and prevents colonization by pathogens. Also found in saliva, tears, and breast milk.	
IgD	1	Functions mainly as an antigen receptor on B cells that have not been exposed to antigens. It has been shown to activate basophils and mast cells to produce antimicrobial factors.	
IgE	1	Binds to allergens and triggers histamine release from mast cells and basophils, and is involved in allergy. Also protects against parasitic worms.	
IgG	4	In its four forms, provides the majority of antibody-based immunity against invading pathogens. The only antibody capable of crossing the placenta to give passive immunity to the fetus.	
IgM	1	Expressed on the surface of B cells (monomer) and in a secreted form (pentamer) with very high avidity. Eliminates pathogens in the early stages of B cell-mediated (humoral) immunity before there is sufficient IgG.	

The co-expression of both of these immunoglobulin isotypes renders the B cell ready to respond to antigen. B cell activation follows engagement of the cell-bound antibody molecule with an antigen, causing the cell to divide and differentiate into an antibody-producing cell called a plasma cell. In this activated form, the B cell starts to produce antibody in a secreted form rather than a membrane-bound form. Some daughter cells of the activated B cells undergo isotype switching, a mechanism that causes the production of antibodies to change from IgM or IgD to the other antibody isotypes, IgE, IgA, or IgG, that have defined roles in the immune system.

Table. Antibody isotypes not found in mammals.

Class	Types	Description
IgY		Found in birds and reptiles; related to mammalian IgG.
IgW		Found in sharks and skates; related to mammalian IgD.

STRUCTURE

Antibodies are heavy (~150 kDa) globular plasma proteins. The size of an antibody molecule is about 10 nm. They have sugar chains (glycans) added to conserved amino acid residues. In other words, antibodies are *glycoproteins*. The attached glycans are critically important to the structure and function of the antibody. Among other things the expressed glycans can modulate an antibody's affinity for its corresponding FcR(s). The basic functional unit of each antibody is an immunoglobulin (Ig) monomer (containing only one Ig unit); secreted antibodies can also be dimeric with two Ig units as with IgA, tetrameric with four Ig units like teleost fish IgM, or pentameric with five Ig units, like mammalian IgM.

The variable parts of an antibody are its V regions, and the constant part is its C region.

Immunoglobulin Domains

The Ig monomer is a “Y”-shaped molecule that consists of four polypeptide chains; two identical *heavy chains* and two identical *light chains* connected by disulfide bonds. Each chain is composed of structural domains called immunoglobulin domains. These domains contain about 70–110 amino acids and are classified into different categories (for example, variable or IgV, and constant or IgC) according to their size and function. They have a characteristic immunoglobulin fold in which two beta sheets create a “sandwich” shape, held together by interactions between conserved cysteines and other charged amino acids.

Heavy Chain

There are five types of mammalian Ig heavy chain denoted by the Greek letters: α , δ , ϵ , γ , and μ . The type of heavy chain present defines the *class* of antibody; these chains are found in IgA, IgD, IgE, IgG, and IgM antibodies, respectively. Distinct heavy chains differ in size and composition; α and γ contain approximately 450 amino acids, whereas μ and ϵ have approximately 550 amino acids. Each heavy chain has two regions, the *constant region* and the *variable region*. The constant region is identical in all antibodies of the same isotype, but differs in antibodies of different isotypes. Heavy chains γ , α and δ have a constant region composed of *three* tandem (in a line) Ig domains, and a hinge region for added flexibility; heavy chains μ and ϵ have a constant region composed of *four* immunoglobulin domains. The variable region of the heavy chain differs in antibodies produced by different B cells, but is the same for all antibodies produced by a single B cell or B cell clone. The variable region of each heavy chain is approximately 110 amino acids long and is composed of a single Ig domain.

Light Chain

In mammals there are two types of immunoglobulin light chain, which are called lambda (λ) and kappa (κ). A light chain has two successive domains: one constant domain and one variable domain. The approximate length of a light chain is 211 to 217 amino acids. Each antibody contains two light chains that are always identical; only one type of light chain, κ or λ , is present per antibody in mammals. Other types of light chains, such as the iota (ι) chain, are found in other vertebrates like sharks (Chondrichthyes) and bony fishes (Teleostei).

CDRs, Fv, Fab and Fc Regions

Some parts of an antibody have the same functions. The arms of the Y, for example, contain the sites that can bind to antigens (in general, identical) and, therefore, recognize specific foreign objects. This region of the antibody is called the *Fab (fragment, antigen-binding) region*. It is composed of one constant and one variable domain from each heavy and light chain of the antibody. The paratope is shaped at the amino terminal end of the antibody monomer by the variable domains from the heavy and light chains. The variable domain is also referred to as the F_V region and is the most important region for binding to antigens. To be specific, variable loops of β -strands, there each on the light (V_L) and heavy (V_H) chains are responsible for binding to the antigen. These loops are referred to as the complementarity determining regions (CDRs). The structures of these CDRs have been clustered and classified by Chothia et al. and more recently by North et al. and Nikoloudis et al. In the framework of the immune network theory, CDRs are also called idiotypes. According to immune network theory, the adaptive immune system is regulated by interactions between idiotypes. The base of the Y plays a role in modulating immune cell activity. This region is called the *Fc (Fragment, crystallizable) region*, and is composed of two heavy chains that contribute two or three constant domains depending on the class of the antibody. Thus, the Fc region ensures that each antibody generates an appropriate immune response for a given antigen, by binding to a specific class of Fc receptors, and other immune molecules, such as complement proteins. By doing this, it mediates different physiological effects including recognition of opsonized particles (binding to $Fc\gamma R$), lysis of cells (binding to complement), and degranulation of mast cells, basophils, and eosinophils (binding to $Fc\epsilon R$).

In summary, the Fab region of the antibody determines antigen specificity while the Fc region of the antibody determines the antibody's class effect. Since only the constant domains of the heavy chains make up the Fc region of an antibody, the classes of heavy chain in antibodies determine their class effects. Possible classes of heavy chains in antibodies include alpha, gamma, delta, epsilon, and mu, and they define the antibody's isotypes IgA, G, D, E, and M, respectively. This infers different isotypes of antibodies have different class effects due to their different Fc regions binding and activating different types of receptors. Possible class effects of antibodies include: Opsonisation, agglutination, haemolysis, complement activation, mast cell degranulation, and neutralisation (though this class effect may be mediated by the Fab region rather than the Fc region). It also implies that Fab-mediated effects are directed at microbes or toxins, whilst Fc mediated effects are directed at effector cells or effector molecules.

FUNCTION

The main categories of antibody action include the following:

- Neutralisation, in which neutralizing antibodies block parts of the surface of a bacterial cell or virion to render its attack ineffective
- Agglutination, in which antibodies “glue together” foreign cells into clumps that are attractive targets for phagocytosis
- Precipitation, in which antibodies “glue together” serum-soluble antigens, forcing them to precipitate out of solution in clumps that are attractive targets for phagocytosis
- Complement activation (fixation), in which antibodies that are latched onto a foreign cell encourage complement to attack it with a membrane attack complex, which leads to the following:
 - (a) Lysis of the foreign cell
 - (b) Encouragement of inflammation by chemotactically attracting inflammatory cells

Activated B cells differentiate into either antibody-producing cells called plasma cells that secrete soluble antibody or memory cells that survive in the body for years afterwards in order to allow the immune system to remember an antigen and respond faster upon future exposures.

At the prenatal and neonatal stages of life, the presence of antibodies is provided by passive immunization from the mother. Early endogenous antibody production varies for different kinds of antibodies, and usually appear within the first years of life. Since antibodies exist freely in the bloodstream, they are said to be part of the humoral immune system. Circulating antibodies are produced by clonal B cells that specifically respond to only one antigen (an example is a virus capsid protein fragment). Antibodies contribute to immunity in three ways: They prevent pathogens from entering or damaging cells by binding to them; they stimulate removal of pathogens by macrophages and other cells by coating the pathogen; and they trigger destruction of pathogens by stimulating other immune responses such as the complement pathway. Antibodies will also trigger vasoactive amine degranulation to contribute to immunity against certain types of antigens (helminths, allergens).

Activation of Complement

Antibodies that bind to surface antigens (for example, on bacteria) will attract the first component of the complement cascade with their Fc region and initiate activation of the “classical” complement system. This results in the killing of bacteria in two ways. First, the binding of the antibody and complement molecules marks the microbe for ingestion by phagocytes in a process called opsonization; these phagocytes are attracted by certain complement molecules generated in the complement cascade. Second, some complement system components form a membrane attack complex to assist antibodies to kill the bacterium directly (bacteriolysis).

Activation of Effector Cells

To combat pathogens that replicate outside cells, antibodies bind to pathogens to link them together, causing them to agglutinate. Since an antibody has at least two paratopes, it can bind more than one antigen by binding identical epitopes carried on the surfaces of these antigens. By coating the pathogen, antibodies stimulate effector functions against the pathogen in cells that recognize their Fc region. Those cells that recognize coated pathogens have Fc receptors, which, as the name suggests, interact with the Fc region of IgA, IgG, and IgE antibodies.

The engagement of a particular antibody with the Fc receptor on a particular cell triggers an effector function of that cell; phagocytes will phagocytose, mast cells and neutrophils will degranulate, natural killer cells will release cytokines and cytotoxic molecules; that will ultimately result in destruction of the invading microbe. The activation of natural killer cells by antibodies initiates a cytotoxic mechanism known as antibody-dependent cell-mediated cytotoxicity (ADCC) – this process may explain the efficacy of monoclonal antibodies used in biological therapies against cancer. The Fc receptors are isotype-specific, which gives greater flexibility to the immune system, invoking only the appropriate immune mechanisms for distinct pathogens.

Natural Antibodies

Humans and higher primates also produce “natural antibodies” that are present in serum before viral infection. Natural antibodies have been defined as antibodies that are produced without any previous infection, vaccination, other foreign antigen exposure or passive immunization. These antibodies can activate the classical complement pathway leading to lysis of enveloped virus particles long before the adaptive immune response is activated. Many natural antibodies are directed against the disaccharide galactose $\alpha(1,3)$ -galactose (α -Gal), which is found as a terminal sugar on glycosylated cell surface proteins, and generated in response to production of this sugar by bacteria contained in the human gut. Rejection of xenotransplanted organs is thought to be, in part, the result of natural antibodies circulating in the serum of the recipient binding to α -Gal antigens expressed on the donor tissue.

IMMUNOGLOBULIN DIVERSITY

Virtually all microbes can trigger an antibody response. Successful recognition and eradication of many different types of microbes requires diversity among antibodies; their amino acid composition varies allowing them to interact with many different antigens. It has been estimated that humans generate about 10 billion different antibodies, each capable of binding a distinct epitope of an antigen. Although a huge repertoire of different antibodies is generated in a single individual, the number of genes available to make these proteins is limited by the size of the human genome. Several complex genetic mechanisms have evolved that allow vertebrate B cells to generate a diverse pool of antibodies from a relatively small number of antibody genes.

Domain Variability

The chromosomal region that encodes an antibody is large and contains several distinct gene loci for each domain of the antibody—the chromosome region containing heavy chain genes (IGH@) is found on chromosome 14, and the loci containing lambda and kappa light chain genes (IGL@ and IGK@) are found on chromosomes 22 and 2 in humans. One of these domains is called the variable domain, which is present in each heavy and light chain of every antibody, but can differ in different antibodies generated from distinct B cells. Differences, between the variable domains, are located on three loops known as hypervariable regions (HV-1, HV-2 and HV-3) or complementarity determining regions (CDR1, CDR2 and CDR3). CDRs are supported within the variable domains by conserved framework regions. The heavy chain locus contains about 65 different variable domain genes that all differ in their CDRs. Combining these genes with an array of genes for other domains of the antibody generates a large cavalry of antibodies with a high degree of variability. This combination is called V(D)J recombination discussed below.

V(D)J Recombination

Somatic recombination of immunoglobulins, also known as *V(D)J recombination*, involves the generation of a unique immunoglobulin variable region. The variable region of each immunoglobulin heavy or light chain is encoded in several pieces—known as gene segments (subgenes). These segments are called variable (V), diversity (D) and joining (J) segments. V, D and J segments are found in Ig heavy chains, but only V and J segments are found in Ig light chains. Multiple copies of the V, D and J gene segments exist, and are tandemly arranged in the genomes of mammals. In the bone marrow, each developing B cell will assemble an immunoglobulin variable region by randomly selecting and combining one V, one D and one J gene segment (or one V and one J segment in the light chain). As there are multiple copies of each type of gene segment, and different combinations of gene segments can be used to generate each immunoglobulin variable region, this process generates a huge number of antibodies, each with different paratopes, and thus different antigen specificities. Interestingly, the rearrangement of several subgenes (*i.e.* V2 family) for lambda light chain immunoglobulin is coupled with the activation of microRNA miR-650, which further influences biology of B-cells. RAG proteins play an important role with V(D)J recombination in cutting DNA at a particular region. Without the presence of these proteins, V(D)J recombination would not occur. After a B cell produces a functional immunoglobulin gene during V(D)J recombination, it cannot express any other variable region (a process known as allelic exclusion) thus each B cell can produce antibodies containing only one kind of variable chain.

Somatic Hypermutation and Affinity Maturation

Following activation with antigen, B cells begin to proliferate rapidly. In these rapidly dividing cells, the genes encoding the variable domains of the heavy and light chains undergo a high rate of point mutation, by a process called *somatic hypermutation* (SHM). SHM results in approximately one nucleotide change per variable

gene, per cell division. As a consequence, any daughter B cells will acquire slight amino acid differences in the variable domains of their antibody chains. This serves to increase the diversity of the antibody pool and impacts the antibody's antigen-binding affinity. Some point mutations will result in the production of antibodies that have a weaker interaction (low affinity) with their antigen than the original antibody, and some mutations will generate antibodies with a stronger interaction (high affinity). B cells that express high affinity antibodies on their surface will receive a strong survival signal during interactions with other cells, whereas those with low affinity antibodies will not, and will die by apoptosis. Thus, B cells expressing antibodies with a higher affinity for the antigen will outcompete those with weaker affinities for function and survival allowing the average affinity of antibodies to increase over time. The process of generating antibodies with increased binding affinities is called *affinity maturation*. Affinity maturation occurs in mature B cells after V(D)J recombination, and is dependent on help from helper T cells.

Class Switching

Isotype or class switching is a biological process occurring after activation of the B cell, which allows the cell to produce different classes of antibody (IgA, IgE, or IgG). The different classes of antibody, and thus effector functions, are defined by the constant (C) regions of the immunoglobulin heavy chain. Initially, naive B cells express only cell-surface IgM and IgD with identical antigen binding regions. Each isotype is adapted for a distinct function; therefore, after activation, an antibody with an IgG, IgA, or IgE effector function might be required to effectively eliminate an antigen. Class switching allows different daughter cells from the same activated B cell to produce antibodies of different isotypes. Only the constant region of the antibody heavy chain changes during class switching; the variable regions, and therefore antigen specificity, remain unchanged. Thus the progeny of a single B cell can produce antibodies, all specific for the same antigen, but with the ability to produce the effector function appropriate for each antigenic challenge. Class switching is triggered by cytokines; the isotype generated depends on which cytokines are present in the B cell environment. Class switching occurs in the heavy chain gene locus by a mechanism called class switch recombination (CSR). This mechanism relies on conserved nucleotide motifs, called *switch (S) regions*, found in DNA upstream of each constant region gene (except in the δ -chain). The DNA strand is broken by the activity of a series of enzymes at two selected S-regions. The variable domain exon is rejoined through a process called non-homologous end joining (NHEJ) to the desired constant region (γ , α or ϵ). This process results in an immunoglobulin gene that encodes an antibody of a different isotype.

Specificity Designations

An antibody can be called *monospecific* if it has specificity for the same antigen or epitope, or *bispecific* if they have affinity for two different antigens or two different epitopes on the same antigen. A group of antibodies can be called *polyvalent* (or *unspecific*) if they have affinity for various antigens or microorganisms. Intravenous immunoglobulin, if not otherwise noted, consists of a variety of different IgG (polyclonal IgG). In contrast, monoclonal antibodies are identical antibodies produced by a single B cell.

Asymmetrical Antibodies

Heterodimeric antibodies, which are also asymmetrical and antibodies, allow for greater flexibility and new formats for attaching a variety of drugs to the antibody arms. One of the general formats for a heterodimeric antibody is the “knobs-into-holes” format. This format is specific to the heavy chain part of the constant region in antibodies. The “knobs” part is engineered by replacing a small amino acid with a larger one. It fits into the “hole”, which is engineered by replacing a large amino acid with a smaller one. What connects the “knobs” to

the “holes” are the disulfide bonds between each chain. The “knobs-into-holes” shape facilitates antibody dependent cell mediated cytotoxicity. Single chain variable fragments (scFv) are connected to the variable domain of the heavy and light chain via a short linker peptide. The linker is rich in glycine, which gives it more flexibility, and serine/threonine, which gives it specificity. Two different scFv fragments can be connected together, via a hinge region, to the constant domain of the heavy chain or the constant domain of the light chain. This gives the antibody bispecificity, allowing for the binding specificities of two different antigens. The “knobs-into-holes” format enhances heterodimer formation but doesn’t suppress homodimer formation. To further improve the function of heterodimeric antibodies, many scientists are looking towards artificial constructs. Artificial antibodies are largely diverse protein motifs that use the functional strategy of the antibody molecule, but aren’t limited by the loop and framework structural constraints of the natural antibody. Being able to control the combinatorial design of the sequence and three-dimensional space could transcend the natural design and allow for the attachment of different combinations of drugs to the arms. Heterodimeric antibodies have a greater range in shapes they can take and the drugs that are attached to the arms don’t have to be the same on each arm, allowing for different combinations of drugs to be used in cancer treatment. Pharmaceuticals are able to produce highly functional bispecific, and even multispecific, antibodies. The degree to which they can function is impressive given that such a change shape from the natural form should lead to decreased functionality.

MEDICAL APPLICATIONS

Disease Diagnosis

Detection of particular antibodies is a very common form of medical diagnostics, and applications such as serology depend on these methods. For example, in biochemical assays for disease diagnosis, a titer of antibodies directed against Epstein-Barr virus or Lyme disease is estimated from the blood. If those antibodies are not present, either the person is not infected or the infection occurred a *very* long time ago, and the B cells generating these specific antibodies have naturally decayed. In clinical immunology, levels of individual classes of immunoglobulins are measured by nephelometry (or turbidimetry) to characterize the antibody profile of patient. Elevations in different classes of immunoglobulins are sometimes useful in determining the cause of liver damage in patients for whom the diagnosis is unclear. For example, elevated IgA indicates alcoholic cirrhosis, elevated IgM indicates viral hepatitis and primary biliary cirrhosis, while IgG is elevated in viral hepatitis, autoimmune hepatitis and cirrhosis. Autoimmune disorders can often be traced to antibodies that bind the body’s own epitopes; many can be detected through blood tests. Antibodies directed against red blood cell surface antigens in immune mediated hemolytic anemia are detected with the Coombs test. The Coombs test is also used for antibody screening in blood transfusion preparation and also for antibody screening in antenatal women. Practically, several immunodiagnostic methods based on detection of complex antigen-antibody are used to diagnose infectious diseases, for example ELISA, immunofluorescence, Western blot, immunodiffusion, immunoelectrophoresis, and magnetic immunoassay. Antibodies raised against human chorionic gonadotropin are used in over the counter pregnancy tests. New dioxaborolane chemistry enables radioactive fluoride (F) labeling of antibodies, which allows for positron emission tomography (PET) imaging of cancer.

Disease Therapy

Targeted monoclonal antibody therapy is employed to treat diseases such as rheumatoid arthritis, multiple sclerosis, psoriasis, and many forms of cancer including non-Hodgkin’s lymphoma, colorectal cancer, head and neck cancer and breast cancer. Some immune deficiencies, such as X-linked agammaglobulinemia and hypogammaglobulinemia, result in partial or complete lack of antibodies. These diseases are often treated by

inducing a short term form of immunity called passive immunity. Passive immunity is achieved through the transfer of ready-made antibodies in the form of human or animal serum, pooled immunoglobulin or monoclonal antibodies, into the affected individual.

Prenatal Therapy

Rh factor, also known as Rh D antigen, is an antigen found on red blood cells; individuals that are Rh-positive (Rh+) have this antigen on their red blood cells and individuals that are Rh-negative (Rh-) do not. During normal childbirth, delivery trauma or complications during pregnancy, blood from a fetus can enter the mother's system. In the case of an Rh-incompatible mother and child, consequential blood mixing may sensitize an Rh- mother to the Rh antigen on the blood cells of the Rh+ child, putting the remainder of the pregnancy, and any subsequent pregnancies, at risk for hemolytic disease of the newborn. Rho(D) immune globulin antibodies are specific for human RhD antigen. Anti-RhD antibodies are administered as part of a prenatal treatment regimen to prevent sensitization that may occur when a Rh-negative mother has a Rh-positive fetus. Treatment of a mother with Anti-RhD antibodies prior to and immediately after trauma and delivery destroys Rh antigen in the mother's system from the fetus. It is important to note that this occurs before the antigen can stimulate maternal B cells to "remember" Rh antigen by generating memory B cells. Therefore, her humoral immune system will not make anti-Rh antibodies, and will not attack the Rh antigens of the current or subsequent babies. Rho(D) Immune Globulin treatment prevents sensitization that can lead to Rh disease, but does not prevent or treat the underlying disease itself.

RESEARCH APPLICATIONS

Specific antibodies are produced by injecting an antigen into a mammal, such as a mouse, rat, rabbit, goat, sheep, or horse for large quantities of antibody. Blood isolated from these animals contains *polyclonal antibodies*—multiple antibodies that bind to the same antigen—in the serum, which can now be called antiserum. Antigens are also injected into chickens for generation of polyclonal antibodies in egg yolk. To obtain antibody that is specific for a single epitope of an antigen, antibody-secreting lymphocytes are isolated from the animal and immortalized by fusing them with a cancer cell line. The fused cells are called hybridomas, and will continually grow and secrete antibody in culture. Single hybridoma cells are isolated by dilution cloning to generate cell clones that all produce the same antibody; these antibodies are called *monoclonal antibodies*. Polyclonal and monoclonal antibodies are often purified using Protein A/G or antigen-affinity chromatography. In research, purified antibodies are used in many applications. Antibodies for research applications can be found directly from antibody suppliers, or through use of a specialist search engine. Research antibodies are most commonly used to identify and locate intracellular and extracellular proteins. Antibodies are used in flow cytometry to differentiate cell types by the proteins they express; different types of cell express different combinations of cluster of differentiation molecules on their surface, and produce different intracellular and secreted proteins. They are also used in immunoprecipitation to separate proteins and anything bound to them (co-immunoprecipitation) from other molecules in a cell lysate, in Western blot analyses to identify proteins separated by electrophoresis, and in immunohistochemistry or immunofluorescence to examine protein expression in tissue sections or to locate proteins within cells with the assistance of a microscope. Proteins can also be detected and quantified with antibodies, using ELISA and ELISPOT techniques.

Antibodies used in research are some of the most powerful, yet most problematic reagents with a tremendous number of factors that must be controlled in any experiment including cross reactivity, or the antibody recognizing multiple epitopes and affinity, which can vary widely depending on experimental conditions such as pH, solvent, state of tissue etc. Multiple attempts have been made to improve both the way that researchers validate antibodies and ways in which they report on antibodies. Researchers using antibodies in their work need to record them

correctly in order to allow their research to be reproducible (and therefore tested, and qualified by other researchers). Less than half of research antibodies referenced in academic papers can be easily identified. Papers published in F1000 in 2014 and 2015 provide researchers with a guide for reporting research antibody use. The RRID paper, is co-published in 4 journals that implemented the RRIDs Standard for research resource citation, which draws data from the antibodyregistry.org as the source of antibody identifiers.

REGULATIONS

Production and Testing

Traditionally, most antibodies are produced by hybridoma cell lines through immortalization of antibody-producing cells by chemically-induced fusion with myeloma cells. In some cases, additional fusions with other lines have created “triomas” and “quadromas”. The manufacturing process should be appropriately described and validated. Validation studies should at least include:

- The demonstration that the process is able to produce in good quality (the process should be validated)
- The efficiency of the antibody purification (all impurities and virus must be eliminated)
- The characterization of purified antibody (physicochemical characterization, immunological properties, biological activities, contaminants,...)
- Determination of the virus clearance studies

Before Clinical Trials

- *Product safety testing:* Sterility (bacteria and fungi), In vitro and in vivo testing for adventitious viruses, Murine retrovirus testing... Product safety data needed before the initiation of feasibility trials in serious or immediately life-threatening conditions, it serves to evaluate dangerous potential of the product.
- *Feasibility testing:* These are pilot studies whose objectives include, among others, early characterization of safety and initial proof of concept in a small specific patient population (in vitro or in vivo testing).

Preclinical Studies

- *Testing cross-reactivity of antibody:* to highlight unwanted interactions (toxicity) of antibodies with previously characterized tissues. This study can be performed in vitro (Reactivity of the antibody or immunoconjugate should be determined with a quick-frozen adult tissues) or in vivo (with appropriate animal models). More informations about in vitro cross-reactivity testing.
- *Preclinical pharmacology and toxicity testing:* Preclinical safety testing of antibody is designed to identify possible toxicities in humans, to estimate the likelihood and severity of potential adverse events in humans, and to identify a safe starting dose and dose escalation, when possible.
- *Animal toxicity studies:* Acute toxicity testing, Repeat-dose toxicity testing, Long-term toxicity testing <http://www.animalresearch.info/en/drug-development/safety-testing/>
- *Pharmacokinetics and pharmacodynamics testing:* Use for determinate clinical dosages, antibody activities (AUC, pharmacodynamics, biodistribution,...), evaluation of the potential clinical effects

STRUCTURE PREDICTION

The importance of antibodies in health care and the biotechnology industry demands knowledge of their structures at high resolution. This information is used for protein engineering, modifying the antigen binding

affinity, and identifying an epitope, of a given antibody. X-ray crystallography is one commonly used method for determining antibody structures. However, crystallizing an antibody is often laborious and time-consuming. Computational approaches provide a cheaper and faster alternative to crystallography, but their results are more equivocal, since they do not produce empirical structures. Online web servers such as *Web Antibody Modeling* (WAM) and *Prediction of Immunoglobulin Structure* (PIGS) enables computational modeling of antibody variable regions. Rosetta Antibody is a novel antibody F_V region structure prediction server, which incorporates sophisticated techniques to minimize CDR loops and optimize the relative orientation of the light and heavy chains, as well as homology models that predict successful docking of antibodies with their unique antigen. The ability to describe the antibody through binding affinity to the antigen is supplemented by information on antibody structure and amino acid sequences for the purpose of patent claims.

IMMUNE COMPLEX

An immune complex, sometimes called an antigen-antibody complex, is a molecule formed from the integral binding of an antibody to a soluble antigen. The bound antigen and antibody act as a unitary object, effectively an antigen of its own with a specific epitope. After an antigen-antibody reaction, the immune complexes can be subject to any of a number of responses, including complement deposition, opsonization, phagocytosis, or processing by proteases. Red blood cells carrying CR1-receptors on their surface may bind C3b-coated immune complexes and transport them to phagocytes, mostly in liver and spleen, and return to the general circulation. Immune complexes may themselves cause illness when they are deposited in organs, for example, in certain forms of vasculitis. This is the third form of hypersensitivity in the Gell-Coombs classification, called type III hypersensitivity. Such hypersensitivity progressing to disease states produces the immune complex diseases. Immune complex deposition is a prominent feature of several autoimmune diseases, including systemic lupus erythematosus, cryoglobulinemia, rheumatoid arthritis, scleroderma and Sjögren's syndrome.

Epidemiology

Epidemiology is the study and analysis of the distribution (who, when, and where) and determinants of health and disease conditions in defined populations. It is the cornerstone of public health, and shapes policy decisions and evidence-based practice by identifying risk factors for disease and targets for preventive health care. Epidemiologists help with study design, collection, and statistical analysis of data, amend interpretation and dissemination of results (including peer review and occasional systematic review). Epidemiology has helped develop methodology used in clinical research, public health studies, and, to a lesser extent, basic research in the biological sciences. Major areas of epidemiological study include disease causation, transmission, outbreak investigation, disease surveillance, forensic epidemiology, occupational epidemiology, screening, biomonitoring, and comparisons of treatment effects such as in clinical trials. Epidemiologists rely on other scientific disciplines like biology to better understand disease processes, statistics to make efficient use of the data and draw appropriate conclusions, social sciences to better understand proximate and distal causes, and engineering for exposure assessment.

HISTORY

The Greek physician Hippocrates is known as the father of medicine, sought a logic to sickness; he is the first person known to have examined the relationships between the occurrence of disease and environmental influences. Hippocrates believed sickness of the human body to be caused by an imbalance of the four humors (air, fire, water and earth “atoms”). The cure to the sickness was to remove or add the humor in question to balance the body. This belief led to the application of bloodletting and dieting in medicine. He coined the terms *endemic* (for diseases usually found in some places but not in others) and *epidemic* (for diseases that are seen at some times but not others).

MODERN ERA

In the middle of the 16th century, a doctor from Verona named Girolamo Fracastoro was the first to propose a theory that these very small, unseeable, particles that cause disease were alive. They were considered to be able to spread by air, multiply by themselves and to be destroyable by fire. In this way he refuted Galen’s miasma theory (poison gas in sick people). In 1543 he wrote a book *De contagione et contagiosis morbis*, in which he was the first to promote personal and environmental hygiene to prevent disease. The development of a sufficiently powerful microscope by Antonie van Leeuwenhoek in 1675 provided visual evidence of living particles consistent with a germ theory of disease. Wu Youke (1582–1652) developed the concept that some diseases were caused by transmissible agents, which he called liqi (pestilential factors). His book *Wenyi Lun* (Treatise on Acute Epidemic Febrile Diseases) can be regarded as the main etiological work that brought forward the concept, ultimately attributed to Westerners, of germs as a cause of epidemic diseases. His concepts are still considered in current scientific research in relation to Traditional Chinese Medicine studies.

Another pioneer, Thomas Sydenham (1624–1689), was the first to distinguish the fevers of Londoners in the later 1600s. His theories on cures of fevers met with much resistance from traditional physicians at the time. He was not able to find the initial cause of the smallpox fever he researched and treated. John Graunt, a haberdasher and amateur statistician, published *Natural and Political Observations... upon the Bills of Mortality* in 1662. In it, he analysed the mortality rolls in London before the Great Plague, presented one of the first life tables, and reported time trends for many diseases, new and old. He provided statistical evidence for many theories on disease, and also refuted some widespread ideas on them. John Snow is famous for his investigations into the causes of the 19th century cholera epidemics, and is also known as the father of (modern) epidemiology. He began with noticing the significantly higher death rates in two areas supplied by Southwark Company. His identification of the Broad Street pump as the cause of the Soho epidemic is considered the classic example of epidemiology. Snow used chlorine in an attempt to clean the water and removed the handle; this ended the outbreak. This has been perceived as a major event in the history of public health and regarded as the founding event of the science of epidemiology, having helped shape public health policies around the world. However, Snow's research and preventive measures to avoid further outbreaks were not fully accepted or put into practice until after his death. Other pioneers include Danish physician Peter Anton Schleisner, who in 1849 related his work on the prevention of the epidemic of neonatal tetanus on the Vestmanna Islands in Iceland. Another important pioneer was Hungarian physician Ignaz Semmelweis, who in 1847 brought down infant mortality at a Vienna hospital by instituting a disinfection procedure. His findings were published in 1850, but his work was ill-received by his colleagues, who discontinued the procedure. Disinfection did not become widely practiced until British surgeon Joseph Lister 'discovered' antiseptics in 1865 in light of the work of Louis Pasteur.

In the early 20th century, mathematical methods were introduced into epidemiology by Ronald Ross, Janet Lane-Claypon, Anderson Gray McKendrick, and others. Another breakthrough was the 1954 publication of the results of a British Doctors Study, led by Richard Doll and Austin Bradford Hill, which lent very strong statistical support to the link between tobacco smoking and lung cancer. In the late 20th century, with advancement of biomedical sciences, a number of molecular markers in blood, other biospecimens and environment were identified as predictors of development or risk of a certain disease. Epidemiology research to examine the relationship between these biomarkers analyzed at the molecular level, and disease was broadly named "molecular epidemiology". Specifically, "genetic epidemiology" has been used for epidemiology of germline genetic variation and disease. Genetic variation is typically determined using DNA from peripheral blood leukocytes. Since the 2000s, genome-wide association studies (GWAS) have been commonly performed to identify genetic risk factors for many diseases and health conditions.

While most molecular epidemiology studies are still using conventional disease diagnosis and classification systems, it is increasingly recognized that disease progression represents inherently heterogeneous processes differing from person to person. Conceptually, each individual has a unique disease process different from any other individual ("the unique disease principle"), considering uniqueness of the exposome (a totality of endogenous and exogenous/ environmental exposures) and its unique influence on molecular pathologic process in each individual. Studies to examine the relationship between an exposure and molecular pathologic signature of disease (particularly cancer) became increasingly common throughout the 2000s. However, the use of molecular pathology in epidemiology posed unique challenges including lack of research guidelines and standardized statistical methodologies, and paucity of interdisciplinary experts and training programmes. Furthermore, the concept of disease heterogeneity appears to conflict with the long-standing premise in epidemiology that individuals with the same disease name have similar etiologies and disease processes. To resolve these issues and advance population health science in the era of molecular precision medicine, "molecular pathology" and "epidemiology" was integrated to create a new interdisciplinary field of "molecular pathological epidemiology" (MPE), defined as "epidemiology of molecular pathology and heterogeneity of disease". In MPE, investigators analyze the relationships between; (A) environmental, dietary, lifestyle and genetic factors; (B) alterations in

cellular or extracellular molecules; and (C) evolution and progression of disease. A better understanding of heterogeneity of disease pathogenesis will further contribute to elucidate etiologies of disease. The MPE approach can be applied to not only neoplastic diseases but also non-neoplastic diseases. The concept and paradigm of MPE have become widespread in the 2010s. By 2012 it was recognized that many pathogens' evolution is rapid enough to be highly relevant to epidemiology, and that therefore much could be gained from an interdisciplinary approach to infectious disease integrating epidemiology and molecular evolution to "inform control strategies, or even patient treatment."

TYPES OF STUDIES

Epidemiologists employ a range of study designs from the observational to experimental and generally categorized as descriptive, analytic (aiming to further examine known associations or hypothesized relationships), and experimental (a term often equated with clinical or community trials of treatments and other interventions). In observational studies, nature is allowed to "take its course," as epidemiologists observe from the sidelines. Conversely, in experimental studies, the epidemiologist is the one in control of all of the factors entering a certain case study. Epidemiological studies are aimed, where possible, at revealing unbiased relationships between exposures such as alcohol or smoking, biological agents, stress, or chemicals to mortality or morbidity. The identification of causal relationships between these exposures and outcomes is an important aspect of epidemiology.

Modern epidemiologists use informatics as a tool. Observational studies have two components, descriptive and analytical. Descriptive observations pertain to the "who, what, where and when of health-related state occurrence". However, analytical observations deal more with the 'how' of a health-related event. Experimental epidemiology contains three case types: randomized controlled trials (often used for new medicine or drug testing), field trials (conducted on those at a high risk of contracting a disease), and community trials (research on social originating diseases). The term 'epidemiologic triad' is used to describe the intersection of *Host*, *Agent*, and *Environment* in analyzing an outbreak.

CASE SERIES

Case-series may refer to the qualitative study of the experience of a single patient, or small group of patients with a similar diagnosis, or to a statistical factor with the potential to produce illness with periods when they are unexposed. The former type of study is purely descriptive and cannot be used to make inferences about the general population of patients with that disease. These types of studies, in which an astute clinician identifies an unusual feature of a disease or a patient's history, may lead to a formulation of a new hypothesis. Using the data from the series, analytic studies could be done to investigate possible causal factors. These can include case-control studies or prospective studies. A case-control study would involve matching comparable controls without the disease to the cases in the series. A prospective study would involve following the case series over time to evaluate the disease's natural history. The latter type, more formally described as self-controlled case-series studies, divide individual patient follow-up time into exposed and unexposed periods and use fixed-effects Poisson regression processes to compare the incidence rate of a given outcome between exposed and unexposed periods. This technique has been extensively used in the study of adverse reactions to vaccination and has been shown in some circumstances to provide statistical power comparable to that available in cohort studies.

CASE-CONTROL STUDIES

Case-control studies select subjects based on their disease status. It is a retrospective study. A group of individuals that are disease positive (the "case" group) is compared with a group of disease negative individuals

(the “control” group). The control group should ideally come from the same population that gave rise to the cases. The case-control study looks back through time at potential exposures that both groups (cases and controls) may have encountered. A 2×2 table is constructed, displaying exposed cases (A), exposed controls (B), unexposed cases (C) and unexposed controls (D). The statistic generated to measure association is the odds ratio (OR), which is the ratio of the odds of exposure in the cases (A/C) to the odds of exposure in the controls (B/D), *i.e.* $OR = (AD/BC)$.

	Cases	Controls
Exposed	A	B
Unexposed	C	D

If the OR is significantly greater than 1, then the conclusion is “those with the disease are more likely to have been exposed,” whereas if it is close to 1 then the exposure and disease are not likely associated. If the OR is far less than one, then this suggests that the exposure is a protective factor in the causation of the disease. Case-control studies are usually faster and more cost effective than cohort studies, but are sensitive to bias (such as recall bias and selection bias). The main challenge is to identify the appropriate control group; the distribution of exposure among the control group should be representative of the distribution in the population that gave rise to the cases.

This can be achieved by drawing a random sample from the original population at risk. This has as a consequence that the control group can contain people with the disease under study when the disease has a high attack rate in a population. A major drawback for case control studies is that, in order to be considered to be statistically significant, the minimum number of cases required at the 95 per cent confidence interval is related to the odds ratio by the equation:

$$\text{total cases} = A + C = 1.96^2 (1 + N) \left(\frac{1}{(\ln(OR))} \right)^2 \left(\frac{OR + 2\sqrt{OR} + 1}{\sqrt{OR}} \right) \approx 15.5 (1 + N) \left(\frac{1}{\ln(OR)} \right)^2$$

where N is the ratio of cases to controls. As the odds ratio approached 1, approaches 0; rendering case control studies all but useless for low odds ratios. For instance, for an odds ratio of 1.5 and cases = controls, the table shown above would look like this:

	Cases	Controls
Exposed	103	84
Unexposed	84	103

For an odds ratio of 1.1:

	Cases	Controls
Exposed	1732	1652
Unexposed	1652	1732

COHORT STUDIES

Cohort studies select subjects based on their exposure status. The study subjects should be at risk of the outcome under investigation at the beginning of the cohort study; this usually means that they should be disease free when the cohort study starts. The cohort is followed through time to assess their later outcome status. An example of a cohort study would be the investigation of a cohort of smokers and non-smokers over time to estimate the incidence of lung cancer. The same 2×2 table is constructed as with the case control study. However, the point estimate generated is the relative risk (RR), which is the probability of disease for a person in the exposed group, $P_e = A / (A + B)$ over the probability of disease for a person in the unexposed group, $P_u = C / (C + D)$, *i.e.* $RR = P_e / P_u$.

.....	Case	Non-case	Total
Exposed	A	B	(A + B)
Unexposed	C	D	(C + D)

As with the OR, a RR greater than 1 shows association, where the conclusion can be read “those with the exposure were more likely to develop disease.” Prospective studies have many benefits over case control studies. The RR is a more powerful effect measure than the OR, as the OR is just an estimation of the RR, since true incidence cannot be calculated in a case control study where subjects are selected based on disease status. Temporality can be established in a prospective study, and confounders are more easily controlled for. However, they are more costly, and there is a greater chance of losing subjects to follow-up based on the long time period over which the cohort is followed. Cohort studies also are limited by the same equation for number of cases as for case control studies, but, if the base incidence rate in the study population is very low, the number of cases required is reduced by $\frac{1}{2}$.

CAUSAL INFERENCE

Although epidemiology is sometimes viewed as a collection of statistical tools used to elucidate the associations of exposures to health outcomes, a deeper understanding of this science is that of discovering *causal* relationships. “Correlation does not imply causation” is a common theme for much of the epidemiological literature. For epidemiologists, the key is in the term inference. Correlation, or at least association between two variables, is a necessary but not sufficient criteria for inference that one variable causes the other. Epidemiologists use gathered data and a broad range of biomedical and psychosocial theories in an iterative way to generate or expand theory, to test hypotheses, and to make educated, informed assertions about which relationships are causal, and about exactly how they are causal. Epidemiologists emphasize that the “one cause – one effect” understanding is a simplistic mis-belief. Most outcomes, whether disease or death, are caused by a chain or web consisting of many component causes. Causes can be distinguished as necessary, sufficient or probabilistic conditions. If a necessary condition can be identified and controlled (*e.g.*, antibodies to a disease agent, energy in an injury), the harmful outcome can be avoided (Robertson, 2015).

BRADFORD HILL CRITERIA

In 1965, Austin Bradford Hill proposed a series of considerations to help assess evidence of causation, which have come to be commonly known as the “Bradford Hill criteria”. In contrast to the explicit intentions of their author, Hill’s considerations are now sometimes taught as a checklist to be implemented for assessing causality. Hill himself said “None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required *sine qua non*.”

- *Strength of Association*: A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.
- *Consistency of Data*: Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.
- *Specificity*: Causation is likely if a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.
- *Temporality*: The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).
- *Biological gradient*: Greater exposure should generally lead to greater incidence of the effect. However, in some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.

- *Plausibility*: A plausible mechanism between cause and effect is helpful (but Hill noted that knowledge of the mechanism is limited by current knowledge).
- *Coherence*: Coherence between epidemiological and laboratory findings increases the likelihood of an effect. However, Hill noted that "... lack of such [laboratory] evidence cannot nullify the epidemiological effect on associations".
- *Experiment*: "Occasionally it is possible to appeal to experimental evidence".
- *Analogy*: The effect of similar factors may be considered.

LEGAL INTERPRETATION

Epidemiological studies can only go to prove that an agent could have caused, but not that it did cause, an effect in any particular case: "Epidemiology is concerned with the incidence of disease in populations and does not address the question of the cause of an individual's disease. This question, sometimes referred to as specific causation, is beyond the domain of the science of epidemiology. Epidemiology has its limits at the point where an inference is made that the relationship between an agent and a disease is causal (general causation) and where the magnitude of excess risk attributed to the agent has been determined; that is, epidemiology addresses whether an agent can cause a disease, not whether an agent did cause a specific plaintiff's disease." In United States law, epidemiology alone cannot prove that a causal association does not exist in general. Conversely, it can be (and is in some circumstances) taken by US courts, in an individual case, to justify an inference that a causal association does exist, based upon a balance of probability. The subdiscipline of forensic epidemiology is directed at the investigation of specific causation of disease or injury in individuals or groups of individuals in instances in which causation is disputed or is unclear, for presentation in legal settings.

POPULATION-BASED HEALTH MANAGEMENT

Epidemiological practice and the results of epidemiological analysis make a significant contribution to emerging population-based health management frameworks.

Population-based health management encompasses the ability to:

- Assess the health states and health needs of a target population;
- Implement and evaluate interventions that are designed to improve the health of that population; and
- Efficiently and effectively provide care for members of that population in a way that is consistent with the community's cultural, policy and health resource values.

Modern population-based health management is complex, requiring a multiple set of skills (medical, political, technological, mathematical etc.) of which epidemiological practice and analysis is a core component, that is unified with management science to provide efficient and effective health care and health guidance to a population. This task requires the forward looking ability of modern risk management approaches that transform health risk factors, incidence, prevalence and mortality statistics (derived from epidemiological analysis) into management metrics that not only guide how a health system responds to current population health issues, but also how a health system can be managed to better respond to future potential population health issues.

Examples of organizations that use population-based health management that leverage the work and results of epidemiological practice include Canadian Strategy for Cancer Control, Health Canada Tobacco Control Programmes, Rick Hansen Foundation, Canadian Tobacco Control Research Initiative. Each of these organizations use a population-based health management framework called Life at Risk that combines epidemiological quantitative analysis with demographics, health agency operational research and economics to perform:

- *Population Life Impacts Simulations*: Measurement of the future potential impact of disease upon the population with respect to new disease cases, prevalence, premature death as well as potential years of life lost from disability and death;

- *Labour Force Life Impacts Simulations*: Measurement of the future potential impact of disease upon the labour force with respect to new disease cases, prevalence, premature death and potential years of life lost from disability and death;
- *Economic Impacts of Disease Simulations*: Measurement of the future potential impact of disease upon private sector disposable income impacts (wages, corporate profits, private health care costs) and public sector disposable income impacts (personal income tax, corporate income tax, consumption taxes, publicly funded health care costs).

APPLIED FIELD EPIDEMIOLOGY

Applied epidemiology is the practice of using epidemiological methods to protect or improve the health of a population. Applied field epidemiology can include investigating communicable and non-communicable disease outbreaks, mortality and morbidity rates, and nutritional status, among other indicators of health, with the purpose of communicating the results to those who can implement appropriate policies or disease control measures.

HUMANITARIAN CONTEXT

As the surveillance and reporting of diseases and other health factors becomes increasingly difficult in humanitarian crisis situations, the methodologies used to report the data are compromised. One study found that less than half (42.4 per cent) of nutrition surveys sampled from humanitarian contexts correctly calculated the prevalence of malnutrition and only one-third (35.3 per cent) of the surveys met the criteria for quality. Among the mortality surveys, only 3.2 per cent met the criteria for quality. As nutritional status and mortality rates help indicate the severity of a crisis, the tracking and reporting of these health factors is crucial. Vital registries are usually the most effective ways to collect data, but in humanitarian contexts these registries can be non-existent, unreliable, or inaccessible. As such, mortality is often inaccurately measured using either prospective demographic surveillance or retrospective mortality surveys. Prospective demographic surveillance requires lots of manpower and is difficult to implement in a spread-out population. Retrospective mortality surveys are prone to selection and reporting biases. Other methods are being developed, but are not common practice yet. WHO, "Health topics: Epidemiology." <http://www.who.int/topics/epidemiology/en/> Accessed: October 30, 2017. Porta, Miquel. "A Dictionary of Epidemiology." Sixth Edition. International Epidemiological Society. Oxford University Press. 2014. Accessed: October 30, 2017. Prudhon, C and Spiegel, P. "A review of methodology and analysis of nutrition and mortality surveys conducted in humanitarian emergencies from October 1993 to April 2004" *Emerging Themes in Epidemiology* 2007, 4:10. <http://www.ete-online.com/content/4/1/10> Accessed: October 30, 2017. Roberts, B et al. "A new method to estimate mortality in crisis-affected and resource-poor settings: validation study." *International Journal of Epidemiology* 2010;39:1584–1596. Accessed: October 30, 2017.

VALIDITY: PRECISION AND BIAS

Different fields in epidemiology have different levels of validity. One way to assess the validity of findings is the ratio of false-positives (claimed effects that are not correct) to false-negatives (studies which fail to support a true effect). To take the field of genetic epidemiology, candidate-gene studies produced over 100 false-positive findings for each false-negative. By contrast genome-wide association appear close to the reverse, with only one false positive for every 100 or more false-negatives. This ratio has improved over time in genetic epidemiology as the field has adopted stringent criteria. By contrast other epidemiological fields have not required such rigorous reporting and are much less reliable as a result.

RANDOM ERROR

Random error is the result of fluctuations around a true value because of sampling variability. Random error is just that: random. It can occur during data collection, coding, transfer, or analysis. Examples of random error include: poorly worded questions, a misunderstanding in interpreting an individual answer from a particular respondent, or a typographical error during coding. Random error affects measurement in a transient, inconsistent manner and it is impossible to correct for random error. There is random error in all sampling procedures. This is called sampling error. Precision in epidemiological variables is a measure of random error. Precision is also inversely related to random error, so that to reduce random error is to increase precision. Confidence intervals are computed to demonstrate the precision of relative risk estimates. The narrower the confidence interval, the more precise the relative risk estimate. There are two basic ways to reduce random error in an epidemiological study. The first is to increase the sample size of the study. In other words, add more subjects to your study. The second is to reduce the variability in measurement in the study. This might be accomplished by using a more precise measuring device or by increasing the number of measurements. Note, that if sample size or number of measurements are increased, or a more precise measuring tool is purchased, the costs of the study are usually increased. There is usually an uneasy balance between the need for adequate precision and the practical issue of study cost.

SYSTEMATIC ERROR

A systematic error or bias occurs when there is a difference between the true value (in the population) and the observed value (in the study) from any cause other than sampling variability. An example of systematic error is if, unknown to you, the pulse oximeter you are using is set incorrectly and adds two points to the true value each time a measurement is taken.

The measuring device could be precise but not accurate. Because the error happens in every instance, it is systematic. Conclusions you draw based on that data will still be incorrect. But the error can be reproduced in the future (*e.g.*, by using the same mis-set instrument).

A mistake in coding that affects *all* responses for that particular question is another example of a systematic error. The validity of a study is dependent on the degree of systematic error. Validity is usually separated into two components:

- Internal validity is dependent on the amount of error in measurements, including exposure, disease, and the associations between these variables. Good internal validity implies a lack of error in measurement and suggests that inferences may be drawn at least as they pertain to the subjects under study.
- External validity pertains to the process of generalizing the findings of the study to the population from which the sample was drawn (or even beyond that population to a more universal statement). This requires an understanding of which conditions are relevant (or irrelevant) to the generalization. Internal validity is clearly a prerequisite for external validity.

Selection Bias

Selection bias occurs when study subjects are selected or become part of the study as a result of a third, unmeasured variable which is associated with both the exposure and outcome of interest. For instance, it has repeatedly been noted that cigarette smokers and non-smokers tend to differ in their study participation rates. (Sackett D cites the example of Seltzer et al., in which 85 per cent of non-smokers and 67 per cent of smokers returned mailed questionnaires.) It is important to note that such a difference in response will not lead to bias if it is not also associated with a systematic difference in outcome between the two response groups.

Information Bias

Information bias is bias arising from systematic error in the assessment of a variable. An example of this is recall bias. A typical example is again provided by Sackett in his discussion of a study examining the effect of specific exposures on fetal health: “in questioning mothers whose recent pregnancies had ended in fetal death or malformation (cases) and a matched group of mothers whose pregnancies ended normally (controls) it was found that 28 per cent of the former, but only 20 per cent of the latter, reported exposure to drugs which could not be substantiated either in earlier prospective interviews or in other health records”. In this example, recall bias probably occurred as a result of women who had had miscarriages having an apparent tendency to better recall and therefore report previous exposures.

Confounding

Confounding has traditionally been defined as bias arising from the co-occurrence or mixing of effects of extraneous factors, referred to as confounders, with the main effect(s) of interest. A more recent definition of confounding invokes the notion of *counterfactual* effects. According to this view, when one observes an outcome of interest, say $Y=1$ (as opposed to $Y=0$), in a given population A which is entirely exposed (*i.e.* exposure $X = 1$ for every unit of the population) the risk of this event will be R_{A1} . The counterfactual or unobserved risk R_{A0} corresponds to the risk which would have been observed if these same individuals had been unexposed (*i.e.* $X = 0$ for every unit of the population). The true effect of exposure therefore is: $R_{A1} - R_{A0}$ (if one is interested in risk differences) or R_{A1}/R_{A0} (if one is interested in relative risk). Since the counterfactual risk R_{A0} is unobservable we approximate it using a second population B and we actually measure the following relations: $R_{A1} - R_{B0}$ or R_{A1}/R_{B0} . In this situation, confounding occurs when $R_{A0} \neq R_{B0}$. (NB: Example assumes binary outcome and exposure variables.) Some epidemiologists prefer to think of confounding separately from common categorizations of bias since, unlike selection and information bias, confounding stems from real causal effects.

THE PROFESSION

To date, few universities offer epidemiology as a course of study at the undergraduate level. One notable undergraduate programme exists at Johns Hopkins University, where students who major in public health can take graduate level courses, including epidemiology, their senior year at the Bloomberg School of Public Health. Although epidemiologic research is conducted by individuals from diverse disciplines, including clinically trained professionals such as physicians, formal training is available through Masters or Doctoral programmes including Master of Public Health (MPH), Master of Science of Epidemiology (MSc.), Doctor of Public Health (DrPH), Doctor of Pharmacy (PharmD), Doctor of Philosophy (PhD), Doctor of Science (ScD). Many other graduate programmes, *e.g.*, Doctor of Social Work (DSW), Doctor of Clinical Practice (DClinP), Doctor of Podiatric Medicine (DPM), Doctor of Veterinary Medicine (DVM), Doctor of Nursing Practice (DNP), Doctor of Physical Therapy (DPT), or for clinically trained physicians, Doctor of Medicine (MD) or Bachelor of Medicine and Surgery (MBBS or MBChB) and Doctor of Osteopathic Medicine (DO), include some training in epidemiologic research or related topics, but this training is generally substantially less than offered in training programmes focused on epidemiology or public health. Reflecting the strong historical tie between epidemiology and medicine, formal training programmes may be set in either schools of public health and medical schools.

As public health/health protection practitioners, epidemiologists work in a number of different settings. Some epidemiologists work ‘in the field’; *i.e.*, in the community, commonly in a public health/health protection service, and are often at the forefront of investigating and combating disease outbreaks. Others work for non-profit organizations, universities, hospitals and larger government entities such as state and local health

departments, various Ministries of Health, Doctors without Borders, the Centers for Disease Control and Prevention (CDC), the Health Protection Agency, the World Health Organization (WHO), or the Public Health Agency of Canada. Epidemiologists can also work in for-profit organizations such as pharmaceutical and medical device companies in groups such as market research or clinical development.

TRANSMISSION (MEDICINE)

In medicine, public health, and biology, transmission is the passing of a pathogen causing communicable disease from an infected host individual or group to a particular individual or group, regardless of whether the other individual was previously infected.

The term strictly refers to the transmission of microorganisms directly from one individual to another by one or more of the following means:

- Droplet contact – coughing or sneezing on another individual
- Direct physical contact – touching an infected individual, including sexual contact
- Indirect physical contact – usually by touching a contaminated surface, including soil (fomite)
- Airborne transmission – if the microorganism can remain in the air for long periods
- Fecal-oral transmission – usually from unwashed hands, contaminated food or water sources due to lack of sanitation and hygiene, an important transmission route in pediatrics, veterinary medicine and developing countries.

Transmission can also be indirect, via another organism, either a vector (*e.g.* a mosquito or fly) or an intermediate host (*e.g.* tapeworm in pigs can be transmitted to humans who ingest improperly cooked pork). Indirect transmission could involve zoonoses or, more typically, larger pathogens like macroparasites with more complex life cycles. Transmissions can be autochthonous (*i.e.* between two individuals in the same place) or may involve travel of the microorganism or the affected hosts.

DEFINITION AND RELATED TERMS

An infectious disease agent can be transmitted in two ways: as horizontal disease agent transmission from one individual to another in the same generation (peers in the same age group) by either direct contact (licking, touching, biting), or indirect contact through air – cough or sneeze (vectors or fomites that allow the transmission of the agent causing the disease without physical contact) or by vertical disease transmission, passing the agent causing the disease from parent to offspring, such as in prenatal or perinatal transmission. The term *infectivity* describes the ability of an organism to enter, survive and multiply in the host, while the *infectiousness* of a disease agent indicates the comparative ease with which the disease agent is transmitted to other hosts. Transmission of pathogen can occur in various ways including physical contact, contaminated food, body fluids, objects, airborne inhalation, or through vector organisms.

ROUTES

The route of transmission is important to epidemiologists because patterns of contact vary between different populations and different groups of populations depending on socio-economic, cultural and other features. For example, low personal and food hygiene due to the lack of a clean water supply may result in increased transmission of diseases by the fecal-oral route, such as cholera. Differences in incidence of such diseases between different groups can also throw light on the routes of transmission of the disease. For example, if it is noted that polio is more common in cities in underdeveloped countries, without a clean water supply, than in cities with a good plumbing system, we might advance the theory that polio is spread by the fecal-oral route.

Airborne

An airborne disease is any disease that is caused by pathogens that can be transmitted through the air. Such diseases include many of considerable importance both in human and veterinary medicine. The relevant pathogens may be viruses, bacteria, or fungi, and they may be spread through breathing, talking, coughing, sneezing, raising of dust, spraying of liquids, toilet flushing or any activities which generates aerosol particles or droplets. Human airborne diseases do not include conditions caused by air pollution such as volatile organic compounds (VOCs), gasses and any airborne particles, though their study and prevention may help inform the science of airborne disease transmission. Airborne diseases include any that are caused via transmission through the air. Many airborne diseases are of great medical importance. The pathogens transmitted may be any kind of microbe, and they may be spread in aerosols, dust or liquids. The aerosols might be generated from sources of infection such as the bodily secretions of an infected animal or person, or biological wastes such as accumulate in lofts, caves, garbage and the like. Such infected aerosols may stay suspended in air currents long enough to travel for considerable distances, though the rate of infection decreases sharply with the distance between the source and the organism infected. Airborne pathogens or allergens often cause inflammation in the nose, throat, sinuses and the lungs. This is caused by the inhalation of these pathogens that affect a person's respiratory system or even the rest of the body. Sinus congestion, coughing and sore throats are examples of inflammation of the upper respiratory air way due to these airborne agents. Air pollution plays a significant role in airborne diseases which is linked to asthma. Pollutants are said to influence lung function by increasing air way inflammation. Many common infections can spread by airborne transmission at least in some cases, including: Anthrax (inhalational), Chickenpox, Influenza, Measles, Smallpox, Cryptococcosis, and Tuberculosis. Airborne diseases can also affect non-humans. For example, Newcastle disease is an avian disease that affects many types of domestic poultry worldwide which is transmitted via airborne contamination. Often, airborne pathogens or allergens cause inflammation in the nose, throat, sinuses, and the upper airway lungs. Upper airway inflammation causes coughing congestion, and sore throat. This is caused by the inhalation of these pathogens that affect a person's respiratory system or even the rest of the body. Sinus congestion, coughing and sore throats are examples of inflammation of the upper respiratory air way due to these airborne agents.

Causes

An airborne disease can be caused by exposure to a source: an infected patient or animal, by being transferred from the infected person or animal's mouth, nose, cut, or needle puncture. People receive the disease through a portal of entry: mouth, nose, cut, or needle puncture.

Transmission

Airborne transmission of disease depends on several physical variables endemic to the infectious particle. Environmental factors influence the efficacy of airborne disease transmission; the most evident environmental conditions are temperature and relative humidity. The sum of all the factors that influence temperature and humidity, either meteorological (outdoor) or human (indoor), as well as other circumstances influencing the spread of the droplets containing the infectious particles, as winds, or human behaviour, sum up the factors influencing the transmission of airborne diseases.

- Climate and living area. Rainfall (number of rainy days being more important than total precipitation), mean of sunshine daily hours, latitude, altitude are characteristic agents to take in account when assessing the possibility of spread of any airborne infection. Furthermore, some infrequent or exceptional extreme events also influence the dissemination of airborne diseases, as tropical storms, hurricanes, typhoons, or monsoons. Climate conditions determine temperature, winds and relative

humidity in any territory, either all year around or at isolated moments (days or weeks). Those are the main factors affecting the spread, duration and infectiousness of droplets containing infectious particles. For instance, influenza virus, is spread easily in northern countries (north hemisphere), because of climate conditions which favour the infectiousness of the virus but on the other hand, in those countries, lots of bacterial infections cannot spread outdoor most of the year, keeping in a latent stage.

UV is harmful to both viruses and bacteria. UV incidence can determine the survival of the infectious particles, so that in those territories with a higher average of sunshine daily hours, and closer to the equator, some particles lose their infectious ability. Infectious particles show an increased survival in the presence of UV light at higher relative humidity levels. It is thought to be due to the protective effect of larger particle sizes, as evaporation would be less at these higher RH levels, showing a protective effect of a thicker water coat.

After isolated events, as tropical storms, has been determined that firstly the quantity of fungal spores is decreased, but a few days later, an exponentially increased number of spores is found, compared to normal conditions.

- Socioeconomics and living conditions. They have a minor role in airborne diseases transmission, but they also have to be taken in consideration. Dwelling is an important aspect. In cities the spread of diseases is faster than in rural areas and outskirts. Normally, cities enclose quarters of buildings, in which the transmission of the viral and bacterial diseases among the neighborhoods is uncomplicated. However, suburban areas are generally more favourable for higher airborne fungal spores

Nearness to large sources of water as rivers and lakes can be a cause of some outbreaks of airborne diseases, after changes in local watershed. Poor sewage systems are usually found in poor countries, especially in the rural areas, and can determine the proliferation of infectious bacteria, that once infecting animal or humans can be transmitted throughout the air. Working conditions, can also settle infectious airborne diseases. At indoor environments, temperature and relative humidity are mainly affected by HVAC systems (heating, ventilation and air conditioning). Inadequate ventilation is implicated in the airborne transmission of respiratory viruses. Poor maintenance or defects on those systems can foster the conditions for airborne infections. For instance, a poor maintenance of air conditioning systems, can lead to an outbreak of *Legionella* (mainly *Legionella pneumophila*), that will spread among the population of the building (workers), before the finding of the focal point. In hospitals, isolation of patients sick of infectious diseases has to be added as a factor, which is noticeable in poor regions, where lack of resources facilitates the spread of infectious diseases.

Prevention

Some ways to prevent airborne diseases include washing hands, using appropriate hand disinfection, getting regular immunizations against diseases believed to be locally present, wearing a respirator and limiting time spent in the presence of any patient likely to be a source of infection. Exposure to a patient or animal with an airborne disease does not guarantee receiving the disease. Because of the changes in host immunity and how much the host was exposed to the particles in the air makes a difference to how the disease affects the body. Antibiotics are not prescribed for patients to control viral infections. They may however be prescribed to a flu patient for instance, to control or prevent bacterial secondary infections. They also may be used in dealing with air-borne bacterial primary infections, such as pneumonic plague. Additionally the Centers for Disease Control and Prevention (CDC) has told consumers about vaccination and following careful hygiene and sanitation protocols for airborne disease prevention. Consumers also have access to preventive measures like UV Air purification devices that FDA and EPA-certified laboratory test data has verified as effective in inactivating a broad array of airborne infectious diseases. Many public health specialists recommend social distancing to reduce the transmission of airborne infections.

Droplet

“Droplet transmission occurs when respiratory droplets generated via coughing, sneezing or talking contact susceptible mucosal surfaces, such as the eyes, nose or mouth. Transmission may also occur indirectly via contact with contaminated fomites with hands and then mucosal surfaces. Respiratory droplets are large and are not able to remain suspended in the air thus they are usually dispersed over short distances.” The pathogen-containing particles, also called *Flügge droplets* (after Carl Flügge), are 0,1–2 mm in diameter, and are reduced by evaporation to *droplet nuclei*—small (smaller than 100 μ in diameter), dry particles that can remain airborne for long periods. Organisms spread by droplet transmission include respiratory viruses (e.g., influenza, parainfluenza virus, adenovirus, respiratory syncytial virus, human metapneumovirus), *Bordetella pertussis*, pneumococci, diphtheria, and rubella.

Fecal–oral

The fecal–oral route (or oral–fecal route or fecal oral route) describes a particular route of transmission of a disease. Pathogens in fecal particles pass from one person to the mouth of another person. Main causes of fecal–oral disease transmission include lack of adequate sanitation (leading to open defecation), and poor hygiene practices. If soil or water bodies are polluted with fecal material, humans can be infected with water-borne diseases or soil-transmitted diseases. Fecal contamination of food is another form of fecal–oral transmission. Washing hands properly after changing a baby’s diaper or going to the bathroom can prevent diseases from spreading. The common fecal–oral routes are fingers, flies, fields, fluids, and food. Analingus, the sexual practice of licking or inserting the tongue into the anus of a partner, is another route. Diseases caused by fecal–oral transmission include diarrhea, typhoid, cholera, polio and hepatitis.

Background

F-Diagram

The foundations for the “F-diagram” being used today were laid down in a publication by WHO in 1958. This publication explained transmission routes and barriers to the transmission of diseases from the focal point of human feces. Modifications have been made over the course of history to derive modern-looking F-diagrams. These diagrams are used in many sanitation publications. They are set up in a way that fecal–oral transmission pathways are shown to take place via water, hands, arthropods and soil. To make it easier to remember, words starting with the letter “F” are used for each of these pathways, namely fluids, fingers, flies, food, fields, fomites (objects and household surfaces).

Rather than only concentrating on human feces, animal feces should also be included in the F-diagram.

The sanitation and hygiene barriers when placed correctly prevent the transmission of an infection through hands, water and food. The F-diagram can be used to show how proper sanitation (in particular toilets, hygiene, handwashing) can act as an effective barrier to stop transmission of diseases via fecal–oral pathways.

Examples

Transmission

The process of transmission may be simple or involve multiple steps. Some examples of routes of fecal–oral transmission include:

- Water that has come in contact with feces (for example due to groundwater pollution from pit latrines) and is then not treated properly before drinking;

- By shaking someone's hand that has been contaminated by stool, changing a child's diapers, working in the garden or dealing with livestock or house pets.
- Food that has been prepared in the presence of fecal matter;
- Disease vectors, like houseflies, spreading contamination from inadequate fecal disposal such as open defecation;
- Poor or absent hand washing after using the toilet or handling feces (such as changing diapers)
- Poor or absent cleaning of anything that has been in contact with feces;
- Sexual practices that may involve oral contact with feces, such as anilingus, coprophilia or "ass to mouth".
- Eating feces, in children, or in a mental disorder called coprophagia
- Eating soil (geophagia)

Prevention

One approach to changing people's behaviors and stopping open defecation is the community-led total sanitation approach. In this process "live demonstrations" of flies moving from food to fresh human feces and back are used. This can "trigger" villagers into action.

Diseases

The list below shows the main diseases that can be passed via the fecal–oral route. They are grouped by the type of pathogen involved in disease transmission.

Bacteria

- *Vibrio cholerae* (cholera)
- *Clostridium difficile* (pseudomembranous enterocolitis)
- *Shigella* (shigellosis/ bacillary dysentery)
- *Salmonella typhi* (typhoid fever)
- *Vibrio parahaemolyticus*
- *Escherichia coli*
- *Campylobacter*

Viruses

- Hepatitis A
- Hepatitis E
- Enteroviruses
- Norovirus acute gastroenteritis
- *Poliovirus* (poliomyelitis)
- Rotavirus – Most of these pathogens cause gastroenteritis.

Protozoans

- *Entameba histolytica* (amoebiasis)
- *Giardia* (giardiasis)
- *Cryptosporidium* (cryptosporidiosis)
- *Toxoplasma gondii* (toxoplasmosis)

Helminths

- Tape worms
- Ascariasis and other soil transmitted helminthiasis

Related Diseases Groupings

Waterborne diseases are diseases caused by pathogenic microorganisms that most commonly are transmitted in contaminated fresh water. This is one particular type of fecal-oral transmission. Neglected tropical diseases also contains many diseases transmitted via the fecal-oral route.

Sexual

Sexually transmitted infections (STI), also referred to as sexually transmitted diseases (STD) or venereal diseases (VD), are infections that are commonly spread by sexual activity, especially vaginal intercourse, anal sex and oral sex. Many times STIs initially do not cause symptoms. This results in a greater risk of passing the disease on to others. Symptoms and signs of disease may include vaginal discharge, penile discharge, ulcers on or around the genitals, and pelvic pain. STIs can be transmitted to an infant before or during childbirth and may result in poor outcomes for the baby. Some STIs may cause problems with the ability to get pregnant. More than 30 different bacteria, viruses, and parasites can be transmitted through sexual activity. Bacterial STIs include chlamydia, gonorrhea, and syphilis. Viral STIs include genital herpes, HIV/AIDS, and genital warts. Parasitic STIs include trichomoniasis. While usually spread by sex, some STIs can be spread by non-sexual contact with donor tissue, blood, breastfeeding, or during childbirth. STI diagnostic tests are usually easily available in the developed world, but this is often not the case in the developing world. The most effective way of preventing STIs is by not having sex. Some vaccinations may also decrease the risk of certain infections including hepatitis B and some types of HPV. Safer sex practices such as use of condoms, having a smaller number of sexual partners, and being in a relationship where each person only has sex with the other also decreases the risk. Circumcision in males may be effective to prevent some infections. During school, comprehensive sex education may also be useful. Most STIs are treatable or curable. Of the most common infections, syphilis, gonorrhea, chlamydia, and trichomoniasis are curable, while herpes, hepatitis B, HIV/AIDS, and HPV are treatable but not curable. Resistance to certain antibiotics is developing among some organisms such as gonorrhea. In 2015, about 1.1 billion people had STIs other than HIV/AIDS. About 500 million were infected with either syphilis, gonorrhea, chlamydia or trichomoniasis. At least an additional 530 million people have genital herpes and 290 million women have human papillomavirus. STIs other than HIV resulted in 108,000 deaths in 2015. In the United States there were 19 million new cases of sexually transmitted infections in 2010. Historical documentation of STIs date back to at least the Ebers papyrus around 1550 BC and the Old Testament. There is often shame and stigma associated with these infections. The term *sexually transmitted infection* is generally preferred over *sexually transmitted disease* or *venereal disease*, as it includes those who do not have symptomatic disease.

Classification

Until the 1990s, STIs were commonly known as *venereal diseases*, the word *venereal* being derived from the Latin word *venereus*, and meaning relating to sexual intercourse or desire, ultimately derived from Venus, the Roman goddess of love. STIs have been euphemistically referred to as “blood diseases” and “social diseases” in the past. The World Health Organization (WHO) has recommended the more inclusive term *sexually transmitted infection* since 1999. Public health officials originally introduced the term *sexually transmitted infection*, which clinicians are increasingly using alongside the term *sexually transmitted disease* in order to distinguish it from the former.

Signs and symptoms

Not all STIs are symptomatic, and symptoms may not appear immediately after infection. In some instances a disease can be carried with no symptoms, which leaves a greater risk of passing the disease on to others. Depending on the disease, some untreated STIs can lead to infertility, chronic pain or even death.

The presence of an STI in prepubescent children may indicate sexual abuse.

Cause

Transmission

A sexually transmitted infection present in a pregnant woman may be passed on to the infant before or after birth. The risks and transmission probabilities of sexually transmitted diseases are summarized by act in the table:

Bacterial

- Chancroid (*Haemophilus ducreyi*)
- Chlamydia (*Chlamydia trachomatis*)
- Gonorrhea (*Neisseria gonorrhoeae*), colloquially known as “the clap”
- Granuloma inguinale or (*Klebsiella granulomatis*)
- *Mycoplasma genitalium*
- *Mycoplasma hominis*
- Syphilis (*Treponema pallidum*)
- Ureaplasma infection

Fungal

- Candidiasis (yeast infection)

Viral

- Viral hepatitis (Hepatitis B virus)—saliva, venereal fluids.
- (Note: Hepatitis A and Hepatitis E are transmitted via the fecal-oral route; Hepatitis C is rarely sexually transmittable, and the route of transmission of Hepatitis D (only if infected with B) is uncertain, but may include sexual transmission.)
- Herpes simplex (Herpes simplex virus 1, 2) skin and mucosal, transmissible with or without visible blisters
- HIV (*Human Immunodeficiency Virus*)—venereal fluids, semen, breast milk, blood
- HPV (*Human Papillomavirus*)—skin and mucosal contact. ‘High risk’ types of HPV cause almost all cervical cancers, as well as some anal, penile, and vulvar cancer. Some other types of HPV cause genital warts.
- Molluscum contagiosum (molluscum contagiosum virus MCV)—close contact

Table. Risk per unprotected sexual act with an infected person.

	Known risks	Possible
Performing oral sex on a man	• Throat chlamydia • Throat gonorrhea (25–30%) • Herpes (rare) • HPV • Syphilis (1%)	• Hepatitis B (low risk) • HIV (0.01%) • Hepatitis C (unknown)
Performing oral sex on a woman	• Herpes • HPV	• Throat gonorrhea • Throat chlamydia
Receiving oral sex—man	• Chlamydia • Gonorrhea • Herpes • Syphilis (1%)	• HPV
Receiving oral sex—woman	• Herpes	• HPV

Receiving oral sex—woman	• Herpes	• HPV • Bacterial vaginosis • Gonorrhea
Vaginal sex—man	• Chlamydia (30–50%) • Crabs • Scabies • Gonorrhea (22%) • Hepatitis B • Herpes (0.07% for HSV-2) • HIV (0.05%) • HPV (high: around 40-50%) • Mycoplasma hominis infection • Mycoplasma genitalium • Syphilis • Trichomoniasis • Ureaplasma infection	• Hepatitis C
Vaginal sex—woman	• Chlamydia (30–50%) • Crabs • Scabies • Gonorrhea (47%) • Hepatitis B (50–70%) • Herpes • HIV (0.1%) • HPV (high; around 40-50%) • Mycoplasma hominis infection • Syphilis • Trichomoniasis • Ureaplasma infection	• Hepatitis C
Anal sex—insertive	• Chlamydia • Crabs • Scabies (40%) • Gonorrhea • Hepatitis B • Herpes • HIV (0.62%) • HPV • Syphilis (14%)	• Hepatitis C
Anal sex—receptive	• Chlamydia • Crabs • Scabies • Gonorrhea • Hepatitis B • Herpes • HIV (1.7%) • HPV • Syphilis (1.4%)	• Hepatitis C
Anilingus	• Amebiasis • Cryptosporidiosis (1%) • Giardiasis • Hepatitis A(1%) • Shigellosis (1%)	• HPV (1%)

Parasites

- Crab louse, colloquially known as “crabs” or “pubic lice” (*Pthirus pubis*) The infestation and accompanying inflammation is Pediculosis pubis
- Scabies (*Sarcoptes scabiei*)
- Trichomoniasis (*Trichomonas vaginalis*), colloquially known as “trich”

Main types

Sexually transmitted infections include:

- Chlamydia is a sexually transmitted infection caused by the bacterium *Chlamydia trachomatis*. In women, symptoms may include abnormal vaginal discharge, burning during urination, and bleeding in between periods, although most women do not experience any symptoms. Symptoms in men include pain when urinating, and abnormal discharge from their penis. If left untreated in both men and women, Chlamydia can infect the urinary tract and potentially lead to pelvic inflammatory disease (PID). PID can cause serious problems during pregnancy and even has the potential to cause infertility. It can cause a woman to have a potentially deadly ectopic pregnancy, in which the egg implants outside of the uterus. However, Chlamydia can be cured with antibiotics.
- The two most common forms of herpes are caused by infection with herpes simplex virus (HSV). HSV-1 is typically acquired orally and causes cold sores, HSV-2 is usually acquired during sexual contact and affects the genitals, however either strain may affect either site. Some people are asymptomatic or have very mild symptoms. Those that do experience symptoms usually notice them 2 to 20 days after exposure which last 2 to 4 weeks. Symptoms can include small fluid-filled blisters, headaches, backaches, itching or tingling sensations in the genital or anal area, pain during urination, Flu like symptoms, swollen glands, or fever. Herpes is spread through skin contact with a person infected with the virus. The virus affects the areas where it entered the body. This can occur through kissing, vaginal intercourse, oral sex or anal sex. The virus is most infectious during times when there are visible symptoms, however those who are asymptomatic can still spread the virus through skin contact. The initial infection and symptoms are usually the most severe because the body does not have any antibodies built up. After the primary attack, one might have recurring attacks that are milder or might not even have future attacks. There is no cure for the disease but there are antiviral medications that treat its symptoms and lower the risk of transmission (Valtrex). Although HSV-1 is typically the “oral” version of the virus, and HSV-2 is typically the “genital” version of the virus, a person with HSV-1 orally CAN transmit that virus to their partner genitally. The virus, either type, will settle into a nerve bundle either at the top of the spine, producing the “oral” outbreak, or a second nerve bundle at the base of the spine, producing the genital outbreak.
- The human papillomavirus (HPV) is the most common STI in the United States. There are more than 40 different strands of HPV and many do not cause any health problems. In 90 per cent of cases the body’s immune system clears the infection naturally within 2 years. Some cases may not be cleared and can lead to genital warts (bumps around the genitals that can be small or large, raised or flat, or shaped like cauliflower) or cervical cancer and other HPV related cancers. Symptoms might not show up until advanced stages. It is important for women to get pap smears in order to check for and treat cancers. There are also two vaccines available for women (Cervarix and Gardasil) that protect against the types of HPV that cause cervical cancer. HPV can be passed through genital-to-genital contact as well as during oral sex. It is important to remember that the infected partner might not have any symptoms.
- Gonorrhea is caused by bacterium that lives on moist mucous membranes in the urethra, vagina, rectum, mouth, throat, and eyes. The infection can spread through contact with the penis, vagina,

mouth or anus. Symptoms of gonorrhea usually appear 2 to 5 days after contact with an infected partner however, some men might not notice symptoms for up to a month. Symptoms in men include burning and pain while urinating, increased urinary frequency, discharge from the penis (white, green, or yellow in colour), red or swollen urethra, swollen or tender testicles, or sore throat. Symptoms in women may include vaginal discharge, burning or itching while urinating, painful sexual intercourse, severe pain in lower abdomen (if infection spreads to fallopian tubes), or fever (if infection spreads to fallopian tubes); however, many women do not show any symptoms. There are some antibiotic resistant strains for Gonorrhea but most cases can be cured with antibiotics.

- Syphilis is an STI caused by a bacterium. Untreated, it can lead to complications and death. Clinical manifestations of syphilis include the ulceration of the uro-genital tract, mouth or rectum; if left untreated the symptoms worsen. In recent years, the prevalence of syphilis has declined in Western Europe, but it has increased in Eastern Europe (former Soviet states). A high incidence of syphilis can be found in places such as Cameroon, Cambodia, Papua New Guinea. Syphilis infections are increasing in the United States.
- Trichomoniasis is a common STI that is caused by infection with a protozoan parasite called *Trichomonas vaginalis*. Trichomoniasis affects both women and men, but symptoms are more common in women. Most patients are treated with an antibiotic called metronidazole, which is very effective.
- HIV (human immunodeficiency virus) damages the body's immune system, which interferes with its ability to fight off disease-causing agents. The virus kills CD4 cells, which are white blood cells that help fight off various infections. HIV is carried in body fluids, and is spread by sexual activity. It can also be spread by contact with infected blood, breast feeding, childbirth, and from mother to child during pregnancy. When HIV is at its most advanced stage, an individual is said to have AIDS (acquired immunodeficiency syndrome). There are different stages of the progression of and HIV infection. The stages include primary infection, asymptomatic infection, symptomatic infection, and AIDS. In the primary infection stage, an individual will have flu like symptoms (headache, fatigue, fever, muscle aches) for about 2 weeks. In the asymptomatic stage, symptoms usually disappear, and the patient can remain asymptomatic for years. When HIV progresses to the symptomatic stage, the immune system is weakened, and has a low cell count of CD4+ T Cells. When the HIV infection becomes life-threatening, it is called AIDS. People with AIDS fall prey to opportunistic infections and die as a result. When the disease was first discovered in the 1980s, those who had AIDS were not likely to live longer than a few years. There are now antiretroviral drugs (ARVs) available to treat HIV infections. There is no known cure for HIV or AIDS but the drugs help suppress the virus. By suppressing the amount of virus in the body, people can lead longer and healthier lives. Even though their virus levels may be low they can still spread the virus to others.

Viruses in Semen

Twenty-seven different viruses have been identified in semen. Information on whether or not transmission occurs or whether the viruses cause disease is uncertain. Some of these microbes are known to be sexually transmitted. Those found in semen are listed by the CDC.

Microbes known to be sexually transmissible (but not generally considered STDs/STIs) include:

- Marburg virus – Virus in semen for seven weeks after clinical recovery.
- HTLV (both types 1 and 2) – Sexually transmissible, consumption of breast milk breastfeeding, and once mistaken as a HIV, risk of leukemia.

Pathophysiology

Many STIs are (more easily) transmitted through the mucous membranes of the penis, vulva, rectum, urinary tract and (less often—depending on type of infection) the mouth, throat, respiratory tract and eyes. The visible membrane covering the head of the penis is a mucous membrane, though it produces no mucus (similar to the lips of the mouth). Mucous membranes differ from skin in that they allow certain pathogens into the body. The amount of contact with infective sources which causes infection varies with each pathogen but in all cases, a disease may result from even light contact from fluid carriers like venereal fluids onto a mucous membrane. Some STIs such as HIV can be transmitted from mother to child either during pregnancy or breastfeeding. Health care professionals suggest safer sex, such as the use of condoms, as a reliable way of decreasing the risk of contracting sexually transmitted diseases during sexual activity, but safer sex cannot be considered to provide complete protection from an STI. The transfer of and exposure to bodily fluids, such as blood transfusions and other blood products, sharing injection needles, needle-stick injuries (when medical staff are inadvertently jabbed or pricked with needles during medical procedures), sharing tattoo needles, and childbirth are other avenues of transmission. These different means put certain groups, such as medical workers, and haemophiliacs and drug users, particularly at risk. It is possible to be an asymptomatic carrier of sexually transmitted diseases. In particular, sexually transmitted diseases in women often cause the serious condition of pelvic inflammatory disease.

Prevention

The Centers for Disease Control and Prevention strategies for reducing STD risk include: vaccination, mutual monogamy, reducing the number of sexual partners and abstinence. The most effective way to prevent sexual transmission of STIs is to avoid contact of body parts or fluids which can lead to transfer with an infected partner. Not all sexual activities involve contact: cybersex, phonesex or masturbation from a distance are methods of avoiding contact. Proper use of condoms reduces contact and risk. Although a condom is effective in limiting exposure, some disease transmission may occur even with a condom. Both partners can get tested for STIs before initiating sexual contact, or before resuming contact if a partner engaged in contact with someone else. Many infections are not detectable immediately after exposure, so enough time must be allowed between possible exposures and testing for the tests to be accurate. Certain STIs, particularly certain persistent viruses like HPV, may be impossible to detect with current medical procedures. Some treatment facilities utilize in-home test kits and have the person return the test for follow-up. Other facilities strongly encourage that those previously infected return to ensure that the infection has been eliminated. Novel strategies to foster re-testing have been the use of text messaging and e-mail as reminders. These types of reminders are now used in addition to phone calls and letters. After obtaining a sexual history, a health care provider can encourage risk reduction by providing prevention counseling. Prevention counseling is most effective if provided in a nonjudgmental and empathetic manner appropriate to the person's culture, language, gender, sexual orientation, age, and developmental level. Prevention counseling for STIs is usually offered to all sexually active adolescents and to all adults who have received a diagnosis, have had an STI in the past year, or have multiple sex partners. USPSTF recommends high-intensity behavioral counseling for all sexually active adolescents and for adults at increased risk for STIs. Such interactive counseling, which can be resource intensive, is directed at a person's risk, the situations in which risk occurs, and the use of personalized goal-setting strategies.

Vaccines

Vaccines are available that protect against some viral STIs, such as Hepatitis A, Hepatitis B, and some types of HPV. Vaccination before initiation of sexual contact is advised to assure maximal protection. The development of vaccines to protect against gonorrhea is ongoing.

Condoms

Condoms and female condoms only provide protection when used properly as a barrier, and only to and from the area that they cover. Uncovered areas are still susceptible to many STIs. In the case of HIV, sexual transmission routes almost always involve the penis, as HIV cannot spread through unbroken skin; therefore, properly shielding the penis with a properly worn condom from the vagina or anus effectively stops HIV transmission. An infected fluid to broken skin borne direct transmission of HIV would not be considered “sexually transmitted”, but can still theoretically occur during sexual contact. This can be avoided simply by not engaging in sexual contact when presenting open, bleeding wounds.

Other STIs, even viral infections, can be prevented with the use of latex, polyurethane or polyisoprene condoms as a barrier. Some microorganisms and viruses are small enough to pass through the pores in natural skin condoms, but are still too large to pass through latex or synthetic condoms.

Proper male condom usage entails:

- Not putting the condom on too tight at the tip by leaving 1.5 centimetres (0.6 in) room for ejaculation. Putting the condom on snug can and often leads to failure.
- Wearing a condom too loose can defeat the barrier
- Avoiding inverting or spilling a condom once worn, whether it has ejaculate in it or not
- If a user attempts to unroll the condom, but realizes they have it on the wrong side, then this condom may not be effective
- Being careful with the condom if handling it with long nails
- Avoiding the use of oil-based lubricants (or anything with oil in it) with latex condoms, as oil can eat holes into them
- Using flavored condoms for oral sex only, as the sugar in the flavoring can lead to yeast infections if used to penetrate

In order to best protect oneself and the partner from STIs, the old condom and its contents are be assumed to be infectious. Therefore, the old condom must be properly disposed of. A new condom is used for each act of intercourse, as multiple usage increases the chance of breakage, defeating the effectiveness as a barrier.

Nonoxynol-9

Researchers had hoped that nonoxynol-9, a vaginal microbicide would help decrease STI risk. Trials, however, have found it ineffective and it may put women at a higher risk of HIV infection.

Screening

Specific age groups, persons who participate in risky sexual behaviour, or those have certain health conditions may require screening. The CDC recommends that sexually active women under the age of 25 and those over 25 at risk should be screened for chlamydia and gonorrhea yearly. Appropriate times for screening are during regular pelvic examinations and preconception evaluations.

Nucleic acid amplification tests are the recommended method of diagnosis for gonorrhea and chlamydia. This can be done on either urine in both men and women, vaginal or cervical swabs in women, or urethral swabs in men. Screening can be performed:

- To assess the presence of infection and prevent tubal infertility in women
- During the initial evaluation before infertility treatment
- To identify HIV infection
- For men who have sex with men
- For those who may have been exposed to hepatitis C
- For HCV

Diagnosis

Testing may be for a single infection, or consist of a number of tests for a range of STIs, including tests for syphilis, trichomonas, gonorrhea, chlamydia, herpes, hepatitis and HIV. No procedure tests for all infectious agents.

STI tests may be used for a number of reasons:

- As a diagnostic test to determine the cause of symptoms or illness
- As a screening test to detect asymptomatic or presymptomatic infections
- As a check that prospective sexual partners are free of disease before they engage in sex without safer sex precautions (for example, when starting a long term mutually monogamous sexual relationship, in fluid bonding, or for procreation).
- As a check prior to or during pregnancy, to prevent harm to the baby
- As a check after birth, to check that the baby has not caught an STI from the mother
- To prevent the use of infected donated blood or organs
- As part of the process of contact tracing from a known infected individual
- As part of mass epidemiological surveillance

Early identification and treatment results in less chance to spread disease, and for some conditions may improve the outcomes of treatment. There is often a window period after initial infection during which an STI test will be negative. During this period, the infection may be transmissible. The duration of this period varies depending on the infection and the test. Diagnosis may also be delayed by reluctance of the infected person to seek a medical professional. One report indicated that people turn to the Internet rather than to a medical professional for information on STIs to a higher degree than for other sexual problems.

Management

In the case of rape, the person can be treated prophylactically with antibiotics. An option for treating partners of patients (index cases) diagnosed with chlamydia or gonorrhea is patient-delivered partner therapy, which is the clinical practice of treating the sex partners of index cases by providing prescriptions or medications to the patient to take to his/her partner without the health care provider first examining the partner.

Epidemiology

In 2008, it was estimated that 500 million people were infected with either syphilis, gonorrhea, chlamydia or trichomoniasis. At least an additional 530 million people have genital herpes and 290 million women have human papillomavirus. STIs other than HIV resulted in 142,000 deaths in 2013. In the United States there were 19 million new cases of sexually transmitted infections in 2010. In 2010, 19 million new cases of sexually transmitted infections occurred in women in the United States. A 2008 CDC study found that 25–40 per cent of U.S. teenage girls has a sexually transmitted disease. AIDS is among the leading causes of death in present-day Sub-Saharan Africa. HIV/AIDS is transmitted primarily via unprotected sexual intercourse. More than 1.1 million persons are living with HIV/AIDS in the United States, and it disproportionately impacts African Americans. Hepatitis B is also considered a sexually transmitted disease because it can be spread through sexual contact. The highest rates are found in Asia and Africa and lower rates are in the Americas and Europe. Approximately two billion people worldwide have been infected with the hepatitis B virus.

History

The first well-recorded European outbreak of what is now known as syphilis occurred in 1494 when it broke out among French troops besieging Naples in the Italian War of 1494–98. The disease may have originated

from the Columbian Exchange. From Naples, the disease swept across Europe, killing more than five million people. As Jared Diamond describes it, “[W]hen syphilis was first definitely recorded in Europe in 1495, its pustules often covered the body from the head to the knees, caused flesh to fall from people’s faces, and led to death within a few months,” rendering it far more fatal than it is today. Diamond concludes, “By 1546, the disease had evolved into the disease with the symptoms so well known to us today.” Gonorrhoeae is recorded at least up to 700 years ago and associated with a district in Paris formerly known as “Le Clapiers”. This is where the prostitutes were to be found at that time. Prior to the invention of modern medicines, sexually transmitted diseases were generally incurable, and treatment was limited to treating the symptoms of the disease. The first voluntary hospital for venereal diseases was founded in 1746 at London Lock Hospital. Treatment was not always voluntary: in the second half of the 19th century, the Contagious Diseases Acts were used to arrest suspected prostitutes. In 1924, a number of states concluded the Brussels Agreement, whereby states agreed to provide free or low-cost medical treatment at ports for merchant seamen with venereal diseases. The first effective treatment for a sexually transmitted disease was salvarsan, a treatment for syphilis. With the discovery of antibiotics, a large number of sexually transmitted diseases became easily curable, and this, combined with effective public health campaigns against STDs, led to a public perception during the 1960s and 1970s that they have ceased to be a serious medical threat. During this period, the importance of contact tracing in treating STIs was recognized. By tracing the sexual partners of infected individuals, testing them for infection, treating the infected and tracing their contacts in turn, STI clinics could effectively suppress infections in the general population. In the 1980s, first genital herpes and then AIDS emerged into the public consciousness as sexually transmitted diseases that could not be cured by modern medicine. AIDS in particular has a long asymptomatic period—during which time HIV (the human immunodeficiency virus, which causes AIDS) can replicate and the disease can be transmitted to others—followed by a symptomatic period, which leads rapidly to death unless treated. HIV/AIDS entered the United States from Haiti in about 1969. Recognition that AIDS threatened a global pandemic led to public information campaigns and the development of treatments that allow AIDS to be managed by suppressing the replication of HIV for as long as possible. Contact tracing continues to be an important measure, even when diseases are incurable, as it helps to contain infection.

Other Names or Slang Terms

In the 1996 sexual harassment case filed against Mitsubishi Motor Manufacturing of America by the Equal Employment Opportunity Commission (EEOC), a number of male employees referred to sexually transmitted infections as “souvenirs”.

Oral

Diseases that are transmitted primarily by oral means may be caught through direct oral contact such as kissing, or by indirect contact such as by sharing a drinking glass or a cigarette. Diseases that are known to be transmissible by kissing or by other direct or indirect oral contact include all of the diseases transmissible by droplet contact and (at least) all forms of herpes viruses, namely Cytomegalovirus infections herpes simplex virus (especially HSV-1) and infectious mononucleosis.

Direct Contact

Diseases that can be transmitted by direct contact are called contagious (contagious is not the same as infectious; although all contagious diseases are infectious, not all infectious diseases are contagious). These diseases can also be transmitted by sharing a towel (where the towel is rubbed vigorously on both bodies) or items of clothing in close contact with the body (socks, for example) if they are not washed thoroughly between

uses. For this reason, contagious diseases often break out in schools, where towels are shared and personal items of clothing accidentally swapped in the changing rooms. Some diseases that are transmissible by direct contact include athlete's foot, impetigo, syphilis (on rare occasions, if an uninfected person touches a chancre), warts, and conjunctivitis.

Vertical

A vertically transmitted infection is an infection caused by pathogens (such as bacteria and viruses) that uses mother-to-child transmission, that is, transmission directly from the mother to an embryo, fetus, or baby during pregnancy or childbirth. It can occur when the mother gets an infection as an intercurrent disease in pregnancy. Nutritional deficiencies may exacerbate the risks of perinatal infection.

Classification

A vertically transmitted infection can be called a perinatal infection if it is transmitted in the perinatal period, which is the period starting at a gestational age of between 22 and 28 weeks (with regional variations in the definition) and ending seven completed days after birth. The term congenital infection can be used if the vertically transmitted infection persists after childbirth.

Examples

Bacteria, viruses, and other organisms are able to be passed from mother to child. Several vertically transmitted infections are included in the TORCH complex, which stands for:

- T – Toxoplasmosis/ *Toxoplasma gondii*
- O – Other infections
- R – Rubella
- C – Cytomegalovirus
- H – Herpes simplex virus-2 or neonatal herpes simplex

“Other infections” include:

- Coxsackievirus
- Chickenpox (caused by varicella zoster virus)
- *Chlamydia*
- HIV

Further information: HIV and pregnancy

- Human T-lymphotropic virus
- Syphilis
- Zika fever, caused by Zika virus, can cause microcephaly and other brain defects in the child.

Hepatitis B may also be classified as a vertically transmitted infection, but the hepatitis B virus is a large virus and does not cross the placenta, hence it cannot infect the fetus unless breaks in the maternal-fetal barrier have occurred, such as can occur in bleeding during childbirth or amniocentesis. The TORCH complex was originally considered to consist of the four conditions mentioned above, with the “TO” referring to *Toxoplasma*. The four-term form is still used in many modern references, and the capitalization “ToRCH” is sometimes used in these contexts. The acronym has also been listed as TORCHES, for TOxoplasmosis, Rubella, Cytomegalovirus, HErpes simplex, and Syphilis.

A further expansion of this acronym, CHEAPTORCHES, was proposed by Ford-Jones and Kellner in 1995:

- C – Chickenpox and shingles
- H – Hepatitis, C (D), E

- E – Enteroviruses
- A – AIDS (HIV infection)
- P – Parvovirus B19 (produces Hydrops foetalis secondary to aplastic anemia)
- T – Toxoplasmosis
- O – Other (Group B Streptococcus, Listeria, Candida, Lyme disease)
- R – Rubella
- C – Cytomegalovirus
- H – Herpes simplex
- E – Everything else sexually transmitted (gonorrhea, *Chlamydia* infection, *Ureaplasma urealyticum*, human papillomavirus)
- S – Syphilis

Signs and Symptoms

The signs and symptoms of a vertically transmitted infection depend on the individual pathogen. It may cause subtle signs such as a influenza-like illness and may not even be noticed by the mother during the pregnancy. In such cases, the effects may be seen first at birth. Symptoms of a vertically transmitted infection may include fever and flu like symptoms. The newborn is often small for gestational age. A petechial rash on the skin may be present, with small reddish or purplish spots due to bleeding from capillaries under the skin. An enlarged liver and spleen (hepatosplenomegaly) is common, as is jaundice. However, jaundice is less common in hepatitis B because a newborn's immune system is not developed well enough to mount a response against liver cells, as would normally be the cause of jaundice in an older child or adult. Hearing impairment, eye problems, mental retardation, autism, and death can be caused by vertically transmitted infections. The mother often has a mild infection with few or no symptoms. The genetic conditions of Aicardi-Goutieres syndrome are possibly present in a similar manner.

Routes

The main routes of transmission of vertically transmitted infections are across the placenta (transplacental) and across the female reproductive tract during childbirth. Transmission is also possible by breaks in the maternal-fetal barrier such by amniocentesis or major trauma.

Transplacental

The embryo and fetus have little or no immune function. They depend on the immune function of their mother. Several pathogens can cross the placenta and cause (perinatal) infection. Often, microorganisms that produce minor illness in the mother are very dangerous for the developing embryo or fetus. This can result in spontaneous abortion or major developmental disorders. For many infections, the baby is more at risk at particular stages of pregnancy. Problems related to perinatal infection are not always directly noticeable.

During Childbirth

Babies can also become infected by their mothers during birth. Some infectious agents may be transmitted to the embryo or fetus in the uterus, while passing through the birth canal, or even shortly after birth. The distinction is important because when transmission is primarily during or after birth, medical intervention can help prevent infections in the infant. During birth, babies are exposed to maternal blood, body fluids, and to the maternal genital tract without the placental barrier intervening. Because of this, blood-borne microorganisms (hepatitis B, HIV), organisms associated with sexually transmitted disease (e.g., *Neisseria gonorrhoeae* and *Chlamydia trachomatis*), and normal fauna of the genitourinary tract (e.g., *Candida albicans*) are among those commonly seen in infection of newborns.

Pathophysiology

Virulence Versus Symbiosis

In the spectrum of optimal virulence, vertical transmission tends to evolve benign symbiosis. It is, therefore, a critical concept for evolutionary medicine. Because a pathogen's ability to pass from parent to child depends significantly on the hosts' ability to reproduce, pathogens' transmissibility tends to be inversely related with their virulence. In other words, as pathogens become more harmful to, and thus decrease the reproduction rate of, their host organism, they are less likely to be passed on to the hosts' offspring, since they will have fewer offspring. Although AIDS is sometimes transmitted through perinatal transmission, its virulence can be accounted for because its primary mode of transmission is not vertical. Moreover, medicine has further decreased the frequency of vertical transmission of AIDS. The incidence of perinatal AIDS cases in the United States has declined as a result of the implementation of recommendations on HIV counselling and voluntary testing practices and the use of zidovudine therapy by providers to reduce perinatal HIV transmission. The price paid in the evolution of symbiosis is, however, great: for many generations, almost all cases of vertical transmission will continue to be pathological—in particular if any other routes of transmission exist. Many generations of random mutation and selection are needed to evolve symbiosis. During this time, the vast majority of vertical transmission cases exhibit the initial virulence. In dual inheritance theory, vertical transmission refers to the passing of cultural traits from parents to children.

Treatment and Prevention

Some vertically transmitted infections, such as toxoplasmosis and syphilis, can be effectively treated with antibiotics if the mother is diagnosed early in her pregnancy. Many viral vertically transmitted infections have no effective treatment, but some, notably rubella and varicella-zoster, can be prevented by vaccinating the mother prior to pregnancy. Pregnant women living in malaria endemic areas are candidates for malaria prophylaxis. It clinically improves the anemia and parasitemia of the pregnant women, and birthweight in their infants. If the mother has active herpes simplex (as may be suggested by a pap test), delivery by Caesarean section can prevent the newborn from contact, and consequent infection, with this virus. IgG₂ antibody may play crucial role in prevention of intrauterine infections and extensive research is going on for developing IgG₂-based therapies for treatment and vaccination.

Prognosis

Each type of vertically transmitted infection has a different prognosis. The stage of the pregnancy at the time of infection also can change the effect on the newborn.

Iatrogenic

Transmission due to medical procedures, such as touching a wound, an injection or transplantation of infected material. Some diseases that can be transmitted iatrogenically include: Creutzfeldt–Jakob disease by injection of contaminated human growth hormone, MRSA and many more.

Vector-borne

In epidemiology, a disease vector is any agent that carries and transmits an infectious pathogen into another living organism; most agents regarded as vectors are organisms, such as intermediate parasites or microbes, but it could be an inanimate medium of infection such as dust particles.

Arthropods

Arthropods form a major group of pathogen vectors with mosquitoes, flies, sand flies, lice, fleas, ticks, and mites transmitting a huge number of pathogens. Many such vectors are haematophagous, which feed on blood at some or all stages of their lives. When the insects blood feed, the pathogen enters the blood stream of the host. This can happen in different ways. The *Anopheles* mosquito, a vector for malaria, filariasis, and various arthropod-borne-viruses (arboviruses), inserts its delicate mouthpart under the skin and feeds on its host's blood. The parasites the mosquito carries are usually located in its salivary glands (used by mosquitoes to anaesthetise the host). Therefore, the parasites are transmitted directly into the host's blood stream. Pool feeders such as the sand fly and black fly, vectors for pathogens causing leishmaniasis and onchocerciasis respectively, will chew a well in the host's skin, forming a small pool of blood from which they feed. *Leishmania* parasites then infect the host through the saliva of the sand fly. *Onchocerca* force their own way out of the insect's head into the pool of blood. Triatomine bugs are responsible for the transmission of a trypanosome, *Trypanosoma cruzi*, which causes Chagas Disease. The Triatomine bugs defecate during feeding and the excrement contains the parasites which are accidentally smeared into the open wound by the host responding to pain and irritation from the bite.

Plants and Fungi

Some plants and fungi act as vectors for various pathogens. For example, the big-vein disease of lettuce was long thought to be caused by a member of the fungal division Chytridiomycota, namely *Olpidium brassicae*. Eventually however, the disease was shown to be viral. Later it transpired that the virus was transmitted by the zoospores of the fungus and also survived in the resting spores. Since then, many other fungi in the Chytridiomycota have been shown to vector plant viruses. Many plant pests that seriously damage important crops depend on other plants, often weeds, to harbour or vector them; the distinction is not always clear. In the case of *Puccinia graminis* for example, *Berberis* and related genera act as alternate hosts in a cycle of infection of grain. More directly, when they twine from one plant to another, parasitic plants such as *Cuscuta* and *Cassytha* have been shown to convey phytoplasmal and viral diseases between plants.

World Health Organization and Vector-borne Disease

The World Health Organization (WHO) states that control and prevention of vector-borne diseases are emphasizing “Integrated Vector Management (IVM)”, which is an approach that looks at the links between health and environment, optimizing benefits to both. In April 2014, WHO launched a campaign called “Small bite, big threat” to educate people about vector-borne illnesses. WHO issued reports indicating that vector-borne illnesses affect poor people, especially people living in areas that do not have adequate levels of sanitation, drinking water and housing.

Vector-borne Zoonotic Disease and Human Activity

Several articles, recent to early 2014, warn that human activities are spreading vector-borne zoonotic diseases. Several articles were published in the medical journal *The Lancet*, and discuss how rapid changes in land use, trade globalization, and “social upheaval” are causing a resurgence in zoonotic disease across the world.

Examples of vector-borne zoonotic diseases include:

- Dengue fever
- Lyme disease
- Plague
- West Nile virus

Many factors affect the incidence of vector-borne diseases. These factors include animals hosting the disease, vectors, and people.

TRACKING

Tracking the transmission of infectious diseases is called disease surveillance. Surveillance of infectious diseases in the public realm traditionally has been the responsibility of public health agencies, either on the (inter)national or a local level. Public health staff rely on health care workers and microbiology laboratories to report cases of reportable diseases to them. The analysis of aggregate data can show the spread of a disease and is at the core of the specialty of epidemiology. To understand the spread of the vast majority of non-notifiable diseases, data either need to be collected in a particular study, or existing data collections can be mined, such as insurance company data or antimicrobial drug sales for example. For diseases transmitted within an institution, such as a hospital, prison, nursing home, boarding school, orphanage, refugee camp etc., infection control specialists are employed, who will review medical records to analyze transmission as part of a hospital epidemiology programme, for example. Because these traditional methods are slow, time-consuming, and labor-intensive, proxies of transmission have been sought. One proxy in the case of influenza is tracking of influenza-like illness at certain sentinel sites of health care practitioners within a state, for example. Tools have been developed to help track influenza epidemics by finding patterns in certain web search query activity. It was found that the frequency of influenza-related web searches as a whole rises as the number of people sick with influenza rises. Examining space-time relationships of web queries has been shown to approximate the spread of influenza and dengue.

Computer simulations of infectious disease spread have been used. Human aggregation can drive transmission, seasonal variation and outbreaks of infectious diseases, such as the annual start of school, bootcamp, the annual Hajj etc. Most recently, data from cell phones have been shown to be able to capture population movements well enough to predict the transmission of certain infectious diseases, like rubella.

RELATIONSHIP WITH VIRULENCE AND SURVIVAL

Pathogens must have a way to be transmitted from one host to another to ensure their species' survival. Infectious agents are generally specialized for a particular method of transmission. Taking an example from the respiratory route, from an evolutionary perspective viruses or bacteria that cause their host to develop coughing and sneezing symptoms have a great survival advantage, as they are much more likely to be ejected from one host and carried to another. This is also the reason that many microorganisms cause diarrhea. The relationship between virulence and transmission is complex, and has important consequences for the long term evolution of a pathogen. Since it takes many generations for a microbe and a new host species to co-evolve, an emerging pathogen may hit its earliest victims especially hard. It is usually in the first wave of a new disease that death rates are highest. If a disease is rapidly fatal, the host may die before the microbe can be passed along to another host. However, this cost may be overwhelmed by the short term benefit of higher infectiousness if transmission is linked to virulence, as it is for instance in the case of cholera (the explosive diarrhea aids the bacterium in finding new hosts) or many respiratory infections (sneezing and coughing create infectious aerosols).

BENEFICIAL MICROORGANISMS

The mode of transmission is also an important aspect of the biology of beneficial microbial symbionts, such as coral-associated dinoflagellates or human microbiota. Organisms can form symbioses with microbes transmitted from their parents, from the environment or unrelated individuals, or both.

Vertical Transmission

Vertical transmission refers to acquisition of symbionts from parents (usually mothers). Vertical transmission can be intracellular (*e.g.* transovarial), or extracellular (for example through post-embryonic contact between parents and offspring). Both intracellular and extracellular vertical transmission can be considered a form of non-genetic inheritance or parental effect. It has been argued that most organisms experience some form of vertical transmission of symbionts. Canonical examples of vertically transmitted symbionts include the nutritional symbiont *Buchnera* in aphids (transovarially transmitted intracellular symbiont) and some components of the human microbiota (transmitted during passage of infants through the birth canal and also through breastfeeding).

Horizontal Transmission

Some beneficial symbionts are acquired horizontally, from the environment or unrelated individuals. This requires that host and symbiont have some method of recognizing each other or each other's products or services. Often, horizontally acquired symbionts are relevant to secondary rather than primary metabolism, for example for use in defence against pathogens, but some primary nutritional symbionts are also horizontally (environmentally) acquired. Additional examples of horizontally transmitted beneficial symbionts include bioluminescent bacteria associated with bobtail squid and nitrogen-fixing bacteria in plants.

Mixed-mode Transmission

Many microbial symbionts, including human microbiota, can be transmitted both vertically and horizontally. Mixed-mode transmission can allow symbionts to have the “best of both worlds” – they can vertically infect host offspring when host density is low, and horizontally infect diverse additional hosts when a number of additional hosts are available. Mixed-mode transmission make the outcome (degree of harm or benefit) of the relationship more difficult to predict, because the evolutionary success of the symbiont is sometimes but not always tied to the success of the host.

CROSS-SPECIES TRANSMISSION

Cross-species transmission, (CST) or spillover, is the ability for a foreign virus, once introduced into an individual of a new host species, to infect that individual and spread throughout a new host population. Steps involved in the transfer of viruses to new hosts include contact between the virus and the host, infection of an initial individual leading to amplification and an outbreak, and the generation within the original or new host of viral variants that have the ability to spread efficiently between individuals in populations of the new host. Often seen in emerging viruses where one species transfers to another, which in turn transfers to humans. Examples include HIV-AIDS, SARS, ebola, swine flu, rabies, and avian influenza. Bacterial pathogens can also be associated with CST. The exact mechanism that facilitates transfer is unknown, however, it is believed that viruses with a rapid mutation rate are able to overcome host-specific immunological defenses. This can occur between species that have high contact rates. It can also occur between species with low contact rates but usually through an intermediary species. Bats, for example, are mammals and can directly transfer rabies to humans through bite and also through aerosolization of bat saliva and urine which are then absorbed by human mucous membranes in the nose, mouth and eyes. A host shifting event is defined as a strain that was previously zoonotic and now circulates exclusively among humans. Similarity between species, for example, transfer between mammals, is believed to be facilitated by similar immunological defenses. Other factors include geographic area, intraspecies behaviours, and phylogenetic relatedness. Virus emergence relies on two factors: initial infection and sustained transmission.

Prevalence

Cross-species transmission is the most significant cause of disease emergence in humans and other species. Wildlife zoonotic diseases of microbial origin are also the most common group of human emerging diseases, and CST between wildlife and livestock has appreciable economic impacts in agriculture by reducing livestock productivity and imposing export restrictions. This makes CST of major concern for public health, agriculture, and wildlife management. A large proportion of viral pathogens that have emerged recently in humans are considered to have originated from various animal species. This is shown by several recent epidemics such as, avian flu, Ebola, monkey pox, and Hanta viruses. There is evidence to suggest that some diseases can potentially be re-introduced to human populations through animal hosts after they have been eradicated in humans. There is a risk of this phenomenon occurring with morbilliviruses as they can readily cross species barriers. CST can also have a significant effect on produce industries. Genotype VI-Avian paramyxovirus serotype 1 (GVI-PMV1) is a virus that arose through cross-species transmission events from Galliformes (*i.e.* chicken) to Columbiformes, and has become prevalent in the poultry industry. CST of rabies virus variants between many different species populations is a major wildlife management concern. Introduction of these variants into non-reservoir animals increases the risk of human exposures and threatens current advances towards rabies control. Many pathogens are thought to have host specialization, which explains the maintenance of distinct strains in host species. Pathogens would have to overcome their host specificity to cross to a new host species. Some studies have argued that host specializations may be exaggerated, and pathogens are more likely to exhibit CST than previously thought. Original hosts usually have low death rates when infected with a pathogen, with fatality rates tending to be much higher in new hosts

Cross-Species Transmission between Humans and Nonhuman Primates

Due to the close relation of humans and nonhuman primates (NHP), disease transmission between NHP and humans is relatively common and can become a major public health concern. Diseases, such as HIV and human adenoviruses have been associated with NHP interactions. In places where contact between humans and NHPs is frequent, precautions are often taken to prevent disease transmission. Simian foamy viruses (SFV) is an enzootic retrovirus that has high rates of cross-species transmission and has been known to affect humans bitten by infected NHPs. It has caused health concerns in places like Indonesia where visitors at monkey temples can contract SFV from temple macaques (*Macaca fascicularis*). TMAdV (titi monkey adenovirus) is a highly divergent, sharing <57 per cent pairwise nucleotide identity with other adenoviruses, NHP virus that had a high fatality rate (83 per cent) in monkeys and is capable of spreading through human hosts.

Predicting and Preventing Transmission between Species

Prediction and monitoring are important for the study of CSTs and their effects. However, factors that determine the origin and fate of cross-species transmission events remain unclear for the majority of human pathogens. This has resulted in the use of different statistical models for the analyzation of CST. Some of these include risk-analysis models, single rate dated tip (SRDT) models, and phylogenetic diffusion models. The study of the genomes of pathogens involved in CST events is very useful in determining their origin and fate. This is because a pathogen's genetic diversity and mutation rate are key factors in determining if it is able to transmit across multiple hosts. This makes it important for the genomes of transmission species to be partially or completely sequenced. A change in genomic structure could cause a pathogen that has narrow host range to become capable of exploiting a wider host range. Genetic distance between different species, geographical range, and other interaction barriers will also influence cross-species transmission. One approach to risk assessment analysis of CST is to develop risk-analysis models that break the “process” of disease transmission

into component parts. Processes and interactions that could lead to cross-species disease transmission are explicitly described as a hypothetical infection chain. Data from laboratory and field experiments is used to estimate the probability of each component, expected natural variation, and margins of error. Different types of CST research would require different analysis pathways to meet their needs. A study on identification of viruses in bats that could spread to other mammals used the workflow: sequencing of genomic samples → “cleaning” of raw reads → elimination of host reads and eukaryotic contaminants → de novo assembly of the remaining reads → annotation of viral contigs → molecular detection of specific viruses → phylogenetic analysis → interpretation of data. Detecting CST and estimating its rate based on prevalence data is challenging. Due to these difficulties computational methods are used to analyse CST events and the pathogens associated with them. The explosive development of molecular techniques has opened new possibilities for using phylogenetic analysis of pathogen genetics to infer epidemiological parameters. This provides some insight into the origins of these events and how they could be addressed. Methods of CST prevention are currently using both biological and computational data. An example of this is using both cellular assays and phylogenetic comparisons to support a role for TRIM5 α , the product of the TRIM5 gene, in suppressing interspecies transmission and emergence of retroviruses in nature.

Cross-Species Transmission Analyzation

Phylogeny

The comparison of genomic data is very important for the study of cross-species transmission. Phylogenetic analysis is used to compare genetic variation in both pathogens associated with CST and the host species that they infect. Taken together, it is possible to infer what allowed a pathogen to crossover to a new host (*i.e.* mutation in a pathogen, change in host susceptibility) and how this can be prevented in the future. If the mechanisms a pathogens uses to initially enter a new species are well characterized and understood a certain level of risk control and prevention can be obtained. In contact, a poor understanding of pathogens, and their associated diseases, makes it harder for preventive measures to be taken. Alternative hosts can also potentially have a critical role in the evolution and diffusion of a pathogen. When a pathogen crosses species it often acquires new characteristics that allow it to breach host barriers. Different pathogen variants can have very different effects on host species. Thus it can be beneficial to CST analysis to compare the same pathogens occurring in different host species. Phylogenetic analysis can be used to track a pathogens history through different species populations. Even if a pathogen is new and highly divergent, phylogenetic comparison can be very insightful. A useful strategy for investigating the history of epidemics caused by pathogen transmission combines molecular clock analysis, to estimate the timescale of the epidemic, and coalescent theory, to infer the demographic history of the pathogen. When constructing phylogenies, computer databases and tools are often used. Programmes, such as BLAST, are used to annotate pathogen sequences, while databases like GenBank provide information about functions based on the pathogens genomic structure. Trees are constructed using computational methods such as MPR or Bayesian Inference, and models are created depending on the needs of the study. Single rate dated tip (SRDT) models, for example, allows for estimates of timescale under a phylogenetic tree. Models for CST prediction will vary depending on what parameters need to be accounted for when constructing the model.

Most Parsimonious Reconstruction (MPR)

Parsimony is the principle in which one chooses the simplest scientific explanation that fits the evidence. In terms of building phylogenetic trees, the best hypothesis is the one that requires the fewest evolutionary changes.

Using parsimony to reconstruct ancestral character states on a phylogenetic tree is a method for testing ecological and evolutionary hypotheses. This method can be used in CST studies to estimate the number of character changes that exist between pathogens in relation to their host. This makes MPR useful for tracking a CST pathogen to its origins. MPR can also be used to compare traits of host species populations. Traits and behaviours within a population could make them more susceptible to CST. For example, species which migrate regionally are important for spreading viruses through population networks. Despite the success of parsimony reconstructions, research suggests they are often sensitive and can sometimes be prone to bias in complex models. This can cause problems for CST models that have to consider many variables. Alternative methods, such as maximum likelihood, have been developed as an alternative to parsimony reconstruction.

Using Genetic Markers

Two methods of measuring genetic variation, variable number tandem repeats (VNTRs) and single nucleotide polymorphisms (SNPs), have been very beneficial to the study of bacterial transmission. VNTRs, due to the low cost and high mutation rates, make them particularly useful to detect genetic differences in recent outbreaks, and while SNPs have a lower mutation rate per locus than VNTRs, they deliver more stable and reliable genetic relationships between isolates. Both methods are used to construct phylogenies for genetic analysis, however, SNPs are more suitable for studies on phylogenies contraction. However, it can be difficult for these methods accurately simulate CST events. Estimates of CST based on phylogenies made using VNTR marker can be biased towards detecting CST events across a wide range of the parameters. SNPs tend to be less biased and variable in estimates of CST when estimations of CST rates are low and low number of SNPs is used.

In general, CST rate estimates using these methods are most reliable in systems with more mutations, more markers, and high genetic differences between introduced strains. CST is very complex and models need to account for a lot of parameters to accurately represent the phenomena. Models that oversimplify reality can result in biased data. Multiple parameters such as number of mutations accumulated since introduction, stochasticity, the genetic difference of strains introduced, and the sampling effort can make unbiased estimates of CST difficult even with whole-genome sequences, especially if sampling is limited, mutation rates are low, or if pathogens were recently introduced. More information on the factors that influence CST rates is needed for the construction of more appropriate models to study these events. The process of using genetic markers to estimate CST rates should take into account several important factors to reduce bias. One, is that the phylogenetic tree constructed in the analysis needs to capture the underlying epidemiological process generating the tree. The models need to account for how the genetic variability of a pathogen influences a disease in a species, not just general differences in genomic structure. Two, the strength of the analysis will depend on the amount of mutation accumulated since the pathogen was introduced in the system. This is due to many models using amount of mutations as an indicator of CST frequency. Therefore, efforts are focused on estimating either time since introduction or the substitution rate of the marker (from laboratory experiments or genomic comparative analysis). This is important not only when using the MPR method but also for Likelihood approaches that require an estimation of the mutation rate. Three, CST will also affect disease prevalence in the potential host, so combining both epidemiological time series data with genetic data may be an excellent approach to CST study.

Bayesian Analysis

Bayesian frameworks are a form of maximum likelihood-based analyses and can be very effective in cross-species transmission studies. Bayesian inference of character evolution methods can account for phylogenetic tree uncertainty and more complex scenarios, with models such as the character diffusion model currently being developed for the study of CST in RNA viruses. A Bayesian statistical approach presents advantages over

other analyses for tracking CST origins. Computational techniques allow integration over an unknown phylogeny, which cannot be directly observed, and unknown migration process, which is usually poorly understood. The Bayesian frameworks are also wellsuited to bring together different kinds of information. The BEAST software, which has a strong focus on calibrated phylogenies and genealogies, illustrates this by offering a large number of complementary evolutionary models including substitution models, demographic and relaxed clock models that can be combined into a full probabilistic model. By adding spatial reconstruction, these models create the probability of biogeographical history reconstruction from genetic data. This could be useful for determining origins of cross-species transmissions. The high effectiveness of Bayesian statistical methods has made them instrumental in evolutionary studies. Bayesian ancestral host reconstruction under discrete diffusion models can be used to infer the origin and effects of pathogens associated with CST. One study on Human adenoviruses using Bayesian supported a gorilla and chimpanzee origin for the viral species, aiding prevention efforts. Despite presumably rare direct contact between sympatric populations of the two species, CST events can occur between them.

The study also determined that two independent HAdV-B transmission events to humans occurred and that the HAdV-Bs circulating in humans are of zoonotic origin and have probably affected global health for most of our species lifetime. Phylogenetic diffusion models are frequently used for phylogeographic analyses, with the inference of host jumping becoming of increasing interest. The Bayesian inference approach enables model averaging over a number of potential diffusion predictors and estimates the support and contribution of each predictor while marginalizing over phylogenetic history. For studying viral CST, the flexibility of the Bayesian statistical framework allows for the reconstruction of virus transmission between different host species while simultaneously testing and quantifying the contribution of multiple ecological and evolutionary influences of both CST spillover and host shifting. One study on rabies in bats showed geographical range overlap is a modest predictor for CST, but not for host shifts. This highlights how Bayesian inferences in models can be used for CST analysis.

EPIDEMIOLOGICAL TRANSITION

In demography and medical geography, epidemiological transition is a phase of development witnessed by a sudden and stark increase in population growth rates brought by improved food security and innovations in public health and medicine, followed by a re-leveling of population growth due to subsequent declines in fertility rates. This accounts for the replacement of infectious diseases by chronic diseases over time due to increased life span as a result of improved health care and disease prevention. This theory was originally posited by Abdel Omran in 1971.

Theory

Omran divided the epidemiological transition of mortality into three phases, in the last of which chronic diseases replace infection as the primary cause of death. These phases are:

- *The Age of Pestilence and Famine:* Mortality is high and fluctuating, precluding sustained population growth, with low and variable life expectancy vacillating between 20 and 40 years. It is characterized by an increase in infectious diseases, malnutrition and famine, common during the Neolithic age. Before the first transition, the hominid ancestors were hunter-gatherers and foragers, a lifestyle partly enabled by a small and dispersed population, however unreliable and seasonal food sources put communities at risk for periods of malnutrition.
- *The Age of Receding Pandemics:* Mortality progressively declines, with the rate of decline accelerating as epidemic peaks decrease in frequency. Average life expectancy increases steadily from about 30 to 50 years. Population growth is sustained and begins to be exponential.

- *The Age of Degenerative and Man-Made Diseases:* Mortality continues to decline and eventually approaches stability at a relatively low level. Mortality is increasingly related to degenerative diseases, cardiovascular disease (CVD), cancer, violence, accidents, and substance abuse, some of these due primarily to human behaviour patterns. The average life expectancy at birth rises gradually until it exceeds 50 years. It is during this stage that fertility becomes the crucial factor in population growth.

In 1998 Barrett et al proposed two additional phases in which cardiovascular diseases diminish as a cause of mortality due to changes in culture, lifestyle and diet, and diseases associated with aging increase in prevalence. In the final phase, disease is largely controlled for those with access to education and health care, but inequalities persist.

- *The Age of Declining CVD Mortality, Aging and Emerging Diseases:* Technological advances in medicine stabilize mortality and the birth rate levels off. Emerging diseases become increasingly lethal due to antibiotic resistance, new pathogens like Ebola or Zika, and mutations that allow old pathogens to overcome human immunity.
- *The Age of Aspired Quality of Life with Persistent Inequalities:* The birth rate declines as lifespan is extended, leading to an age-balanced population. Socioeconomic, ethnic, and gender inequalities continue to manifest differences in mortality and fertility.

The epidemiological transition occurs when a country undergoes the process of modernization from developing nation to developed nation status. The developments of modern health care and medicine, such as antibiotics, drastically reduce infant mortality rates and extend average life expectancy which, coupled with subsequent declines in fertility rates, reflects a transition to chronic and degenerative diseases as more important causes of death. The theory of epidemiological transition uses patterns of health and disease as well as their forms of demographic, economical and sociological determinants and outcomes.

History

In general human history, Omran's first phase occurs when human population sustains cyclic, low-growth, and mostly linear, up-and-down patterns associated with wars, famine, epidemic outbreaks, as well as small golden ages, and localized periods of "prosperity". In early pre-agricultural history, infant mortality rates were high and average life expectancy low. Today, life expectancy in developing countries remains relatively low, as in many Sub-Saharan African nations where it typically doesn't exceed 60 years of age. The second phase involves improved nutrition as a result of stable food production along with advances in medicine and the development of health care systems. Mortality in Western Europe and North America was halved during the 19th century due to closed sewage systems and clean water provided by public utilities, with a particular benefit for children of both sexes and to females in the adolescent and reproductive age periods, probably because the susceptibility of these groups to infectious and deficiency diseases is relatively high. An overall reduction in malnutrition enabled populations to better resist infectious disease. Treatment breakthroughs of importance included the initiation of vaccination during the early nineteenth century, and the discovery of penicillin in the mid 20th century, which led respectively to a widespread and dramatic decline in death rates from previously serious diseases such as smallpox and sepsis. Population growth rates surged in the 1950s, 1960's and 1970's to 1.8 per cent per year and higher, with the world gaining 2 billion people between 1950 and the 1980s. A decline in mortality without a corresponding decline in fertility leads to a population pyramid assuming the shape of a bullet or a barrel, as young and middle-age groups comprise equivalent percentages of the population.

Omran's third phase occurs when human birth rates drastically decline from highly positive replacement rates to stable replacement numbers. In several European nations replacement rates have even become negative. This transition generally represents the net effect of individual choices on family size and the ability to implement those choices. Omran gives three possible factors tending to encourage reduced fertility rates:

- *Bio-physiologic factors*, associated with reduced infant mortality and the expectation of longer life in parents;
- *Socioeconomic factors*, associated with childhood survival and the economic challenges of large family size; and
- *Psychological or emotional factors*, where society as a whole changes its rationale and opinion on family size and parental energies are redirected to qualitative aspects of child-raising.

Impact on Fertility

Improvements in female and childhood survival that occur with the shift in health and disease patterns discussed above have distinct and seemingly contradictory effects on fertility. While better health and greater longevity enjoyed by females of reproductive age tend to enhance fertility, the reduced risks to infants and young children that occurs in the later stages of the transition tends to have the opposite effect: prolonged breastfeeding associated with reduced mortality among infants and toddlers, together with parental recognition of improved childhood survival, tend to lengthen birth intervals and depress overall reproductive rates.

Economic Impact

The transition may also be associated with demographic movements to urban areas, and a shift from agriculture and labor-based production output to technological and service-sector-based economies. This shift in demographic and disease profiles is currently under way in most developing nations, however every country is unique and transition speed is based on numerous geographical and sociopolitical factors. Whether the transition is due to socioeconomic improvements (as in developed countries) or by modern public health programmes (as has been the case in many developing countries), the lowering of mortality and of infectious disease tends to increase economic productivity through better functioning of adult members of the labour force and through an increase in the proportion of children who survive and mature into productive members of society.

Models of Transition

Omran developed three models to explain the epidemiological transition.

- *Classical/Western model*: (England, Wales, and Sweden) Countries in Western Europe typically experienced a transition that began in the late eighteenth century and lasted over 150 years to the post-World War II era. The lengthy transition allowed fertility to decline at virtually the same rate that mortality also declined. Germany might be considered another example of this model.
- *Accelerated model*: (Japan) Japan experienced a rapid transition as a result of a few decades of intensive war-driven industrialization followed by postwar occupation. The accelerated transition follows a pattern similar to the Classical/Western Model except that it occurs within a much shorter time span. China might be considered another example of this model.
- *Contemporary/Delayed model*: (Chile, Ceylon) Due to slow economic development, Chile and Ceylon (Sri Lanka) experienced delayed transitions that have lasted into the 21st century. Medical and public health improvements have reduced mortality, while the birth rate continues to remain high. Cultural traditions combined with political and economic instability and food insecurity mean that mortality for women and children fluctuates more than for men. Mauritius might be considered another example of this model.

Determinants of Disease

- *Ecobiological*: changing patterns of immunity, vectors (such as the black rat partially responsible for spreading bubonic plague in Europe), and the movement of pathogenic organisms. These alter the frequency of epidemic infectious diseases as well as chronic infections and other illnesses that affect fertility and infant mortality.
- *Socioeconomic*: political and cultural determinants, including standards of living, health habits, hygiene and nutrition. Hygiene and nutrition are included here, rather than under medical determinants, because their improvement in western countries was largely a byproduct of social change rather than a result of medical design.
- *Medical/Public health*: specific preventive and curative measures used to combat disease, including improved public sanitation, immunization and the development of decisive therapies. Medical and public health factors came into play late in the western transition, but have an influence early in certain accelerated and contemporary transitions.

Other Perspectives

McMichael, Preston, and Murray offer a more nuanced view of the epidemiological transition, highlighting macro trends and emphasizing that there is a change from infectious to non-communicable disease, but arguing that it happens differently in different contexts. One of the first to refine the idea of the epidemiological transition was Preston, who in 1976 proposed the first comprehensive statistical model relating mortality and cause-specific mortality. Preston used life tables from 43 national populations, including both developed countries such as United States and England and developing countries such as Chile, Colombia, Costa Rica, Guatemala, México, Panama, Taiwan, Trinidad and Tobago, and Venezuela. He used multiple linear regression to analyze the cause-specific-age-standardized death rates by sex.

The estimated slopes represented the proportional contribution of each cause to a unit change in the total mortality rate. With the exception of neoplasms in both sexes and cardiovascular disease in males, all of the estimated slopes were positive and statistically significant. This demonstrated that the mortality rates from each specific cause were expected to decline as total mortality declined. The major causes accounting for the decline were all infectious and parasitic diseases.

McMichael et al. argue (2004) that the epidemiological transition has not taken place homogeneously in all countries. Countries have varied in the speed with which they go through the transition as well as what stage of the transition they are in. The global burden of disease web site provides visual comparisons of the disease burdens of countries and the changes over time. The epidemiological transition correlates with changes in life expectancy. Worldwide, mortality rates have decreased as both technological and medical advancements have led to a tremendous decrease in infectious diseases. With fewer people dying from infectious diseases, there is a rising prevalence of chronic and/or degenerative diseases in the older surviving population.

McMichael et al. describe life expectancy trends as grouped into three categories, as suggested by Casselli et al.:

- Rapid gains among countries such as Chile, Mexico and Tunisia that have strong economic and technical relationships with developed countries
- Slower plateauing gains mostly among developed countries with slower increases in life expectancy (for example, France)
- Frank reversals occurring mostly in developing countries where the HIV epidemic led to a significant decline in life expectancy, and countries in the former Soviet Union, afflicted by social upheavals, heavy alcohol consumption and institutional inadequacy (for example, Zimbabwe and Botswana)

Murray and Lopez (1996) offered one of the most important cause-of-death models as part of the 1990 Global Burden of Disease Study. Their “cause of death” patterns sought to describe the fraction of deaths attributed to a set of mutually exclusive and collectively exhaustive causes. They divided diseases into three cause groups and made several important observations:

- Group 1 - communicable, maternal, perinatal, and nutritional: These causes of death decline much faster than overall mortality and comprise a small fraction of deaths in wealthier countries.
- Group 2 - non-communicable diseases: These causes of death are a major challenge for countries that have completed or nearly completed the epidemiological transition.
- Group 3 - injuries: This cause of death is most variable within and across different countries and is less predictive of all-cause mortality.

The regression approach underlying the Global Burden of Disease received some critique in light of real-world violations of the model’s “mutually exclusive and collectively exhaustive” cause attribution. Building on the existing body of evidence, Salomon and Murray (2002), further add nuances to the traditional theory of epidemiological transition by disintegrating it based on disease categories and different age-sex groups, positing that the epidemiological transition entails a real transition in the cause composition of age-specific mortality, as opposed to just a transition in the age structure. Using Global Burden of Disease data from 1990, they disintegrate the transition across three cause groups: communicable diseases, non-communicable diseases and injuries, seeking to explain the variation in all-cause mortality as a function of cause-specific mortality in 58 countries from 1950 to 1998. This analysis validates the underlying premise of the classic epidemiological transition theory: as total mortality declines and income rises, communicable diseases cause less and less mortality compared to non-communicable diseases and injuries. Decomposing this overall impact by age-sex groups, they find that for males, when overall mortality decreases, the importance of non-communicable diseases (NCDs) increases relative to the other causes with an age-specific impact on the role of injuries, whereas for women, both NCDs and injuries gain a more significant share with mortality decreases. For children over one year, they find that there is a gradual transition from communicable to non-communicable diseases, with injuries remaining significant in males. For young adults, the epidemiological transition is particularly different: for males, there is a shift from injuries to NCDs in lower income settings, and the opposite in higher-income settings; for females, rising income also signifies a shift from NCDs to injuries, but the role of injuries becomes more significant over time compared to males. Finally, for both males and females over 50, there is no epidemiological transition impact on the cause composition of mortality.

Current Evidence

The majority of the literature on the epidemiological transition that was published since these seminal papers confirms the context-specific nature of the epidemiological transition: while there is an overall all-cause mortality decline, the nature of cause-specific mortality declines differs across contexts. Increasing obesity rates in high-income countries are further confirming the epidemiological transition theory as the epidemic leads to an increase in NCDs. The picture is more nuanced in low- and middle-income countries, where there are signs of a protracted transition with the double burden of communicable and noncommunicable disease. A recent review of cause-specific mortality rates from 12 low- and middle-income countries in Asia and sub-Saharan Africa by Santosa and Byass (2016) shows that broadly, low- and middle-income countries are rapidly transitioning to lower total mortality and lower infectious disease mortality. A more macro-level analysis from the Global Burden of Disease data conducted by Murray and others (2015) finds that while there is a global trend towards decreasing mortality and increasing NCD prevalence, this global trend is being driven by country-specific effects as opposed to a broader transition; further, there are varying patterns within and between countries, which makes it difficult to have a single unified theory of epidemiological transition.

A theory of epidemiological transition aimed at explaining not just describing changes in population disease and mortality profiles would need to encompass the role in different morbid conditions of infectious diseases contracted over the life course. The concept of linear transition from infectious diseases to other conditions referred to as degenerative or non-communicable, was based on a false dichotomy as common microorganisms have now been confirmed as causal agents in several conditions recorded as the underlying cause of many deaths. A revised transition model might focus more on disease aetiology and the determinants of cause-specific mortality change, while encompassing the possibility that infectious causation may be established for other morbid conditions through the vast amount of ongoing research into associations with infectious diseases.

Medical Diagnosis

INTRODUCTION

Medical diagnosis (abbreviated Dx or D_S) is the process of determining which disease or condition explains a person's symptoms and signs. It is most often referred to as diagnosis with the medical context being implicit. The information required for diagnosis is typically collected from a history and physical examination of the person seeking medical care. Often, one or more diagnostic procedures, such as diagnostic tests, are also done during the process. Sometimes posthumous diagnosis is considered a kind of medical diagnosis. Diagnosis is often challenging, because many signs and symptoms are nonspecific. For example, redness of the skin (erythema), by itself, is a sign of many disorders and thus does not tell the health care professional what is wrong. Thus differential diagnosis, in which several possible explanations are compared and contrasted, must be performed. This involves the correlation of various pieces of information followed by the recognition and differentiation of patterns. Occasionally the process is made easy by a sign or symptom (or a group of several) that is pathognomonic. Diagnosis is a major component of the procedure of a doctor's visit. From the point of view of statistics, the diagnostic procedure involves classification tests.

HISTORY

The first recorded examples of medical diagnosis are found in the writings of Imhotep (2630–2611 BC) in ancient Egypt (the Edwin Smith Papyrus). A Babylonian medical textbook, the *Diagnostic Handbook* written by Esagil-kin-apli (fl.1069–1046 BC), introduced the use of empiricism, logic and rationality in the diagnosis of an illness or disease. Traditional Chinese Medicine, as described in the Yellow Emperor's Inner Canon or Huangdi Neijing, specified four diagnostic methods: inspection, auscultation-olfaction, interrogation, and palpation. Hippocrates was known to make diagnoses by tasting his patients' urine and smelling their sweat.

MEDICAL USES

A diagnosis, in the sense of diagnostic procedure, can be regarded as an attempt at classification of an individual's condition into separate and distinct categories that allow medical decisions about treatment and prognosis to be made. Subsequently, a diagnostic opinion is often described in terms of a disease or other condition, but in the case of a wrong diagnosis, the individual's actual disease or condition is not the same as the individual's diagnosis.

A diagnostic procedure may be performed by various health care professionals such as a physician, physical therapist, optometrist, health care scientist, chiropractor, dentist, podiatrist, nurse practitioner, or physician assistant. This article uses *diagnostician* as any of these person categories. A diagnostic procedure (as well as the opinion reached thereby) does not necessarily involve elucidation of the etiology of the diseases or

conditions of interest, that is, what *caused* the disease or condition. Such elucidation can be useful to optimize treatment, further specify the prognosis or prevent recurrence of the disease or condition in the future.

The initial task is to detect a medical indication to perform a diagnostic procedure. Indications include:

- Detection of any deviation from what is known to be normal, such as can be described in terms of, for example, anatomy (the structure of the human body), physiology (how the body works), pathology (what can go wrong with the anatomy and physiology), psychology (thought and behaviour) and human homeostasis (regarding mechanisms to keep body systems in balance). Knowledge of what is normal and measuring of the patient's current condition against those norms can assist in determining the patient's particular departure from homeostasis and the degree of departure, which in turn can assist in quantifying the indication for further diagnostic processing.
- A complaint expressed by a patient.
- The fact that a patient has sought a diagnostician can itself be an indication to perform a diagnostic procedure. For example, in a doctor's visit, the physician may already start performing a diagnostic procedure by watching the gait of the patient from the waiting room to the doctor's office even before she or he has started to present any complaints.

Even during an already ongoing diagnostic procedure, there can be an indication to perform another, separate, diagnostic procedure for another, potentially concomitant, disease or condition. This may occur as a result of an incidental finding of a sign unrelated to the parameter of interest, such as can occur in comprehensive tests such as radiological studies like magnetic resonance imaging or blood test panels that also include blood tests that are not relevant for the ongoing diagnosis.

PROCEDURE

General components which are present in a diagnostic procedure in most of the various available methods include:

- Complementing the already given information with further data gathering, which may include questions of the medical history (potentially from other people close to the patient as well), physical examination and various diagnostic tests. A diagnostic test is any kind of medical test performed to aid in the diagnosis or detection of disease. Diagnostic tests can also be used to provide prognostic information on people with established disease.
- Processing of the answers, findings or other results. Consultations with other providers and specialists in the field may be sought.

There are a number of methods or techniques that can be used in a diagnostic procedure, including performing a differential diagnosis or following medical algorithms. In reality, a diagnostic procedure may involve components of multiple methods.

Differential Diagnosis

The method of differential diagnosis is based on finding as many candidate diseases or conditions as possible that can possibly cause the signs or symptoms, followed by a process of elimination or at least of rendering the entries more or less probable by further medical tests and other processing until, aiming to reach the point where only one candidate disease or condition remains as probable. The final result may also remain a list of possible conditions, ranked in order of probability or severity.

The resultant diagnostic opinion by this method can be regarded more or less as a diagnosis of exclusion. Even if it does not result in a single probable disease or condition, it can at least rule out any imminently life-threatening conditions. Unless the provider is certain of the condition present, further medical tests, such as medical imaging, are performed or scheduled in part to confirm or disprove the diagnosis but also to document

the patient's status and keep the patient's medical history up to date. If unexpected findings are made during this process, the initial hypothesis may be ruled out and the provider must then consider other hypotheses.

Pattern Recognition

In a pattern recognition method the provider uses experience to recognize a pattern of clinical characteristics. It is mainly based on certain symptoms or signs being associated with certain diseases or conditions, not necessarily involving the more cognitive processing involved in a differential diagnosis. This may be the primary method used in cases where diseases are “obvious”, or the provider's experience may enable him or her to recognize the condition quickly. Theoretically, a certain pattern of signs or symptoms can be directly associated with a certain therapy, even without a definite decision regarding what is the actual disease, but such a compromise carries a substantial risk of missing a diagnosis which actually has a different therapy so it may be limited to cases where no diagnosis can be made.

Diagnostic Criteria

The term *diagnostic criteria* designates the specific combination of signs, symptoms, and test results that the clinician uses to attempt to determine the correct diagnosis.

Some examples of diagnostic criteria, also known as clinical case definitions, are:

- Amsterdam criteria for hereditary nonpolyposis colorectal cancer
- McDonald criteria for multiple sclerosis
- ACR criteria for systemic lupus erythematosus
- Centor criteria for strep throat

Clinical Decision Support System

Clinical decision support systems are interactive computer programmes designed to assist health professionals with decision-making tasks. The clinician interacts with the software utilizing both the clinician's knowledge and the software to make a better analysis of the patients data than either human or software could make on their own. Typically the system makes suggestions for the clinician to look through and the clinician picks useful information and removes erroneous suggestions.

Other Diagnostic Procedure Methods

Other methods that can be used in performing a diagnostic procedure include:

- Usage of medical algorithms
- An “exhaustive method”, in which every possible question is asked and all possible data is collected. This is often referred to as a “diagnostic workup”.
- Use of a sensory pill that collects and transmits physiological information after being swallowed.
- Using optical coherence tomography to produce detailed images of the brain or other soft tissue, through a “window” made of zirconia that has been modified to be transparent and implanted in the skull.

ADVERSE EFFECTS

Diagnosis problems are the dominant cause of medical malpractice payments, accounting for 35 per cent of total payments in a study of 25 years of data and 350,000 claims.

Overdiagnosis

Overdiagnosis is the diagnosis of “disease” that will never cause symptoms or death during a patient’s lifetime. It is a problem because it turns people into patients unnecessarily and because it can lead to economic waste (overutilization) and treatments that may cause harm. Overdiagnosis occurs when a disease is diagnosed correctly, but the diagnosis is irrelevant. A correct diagnosis may be irrelevant because treatment for the disease is not available, not needed, or not wanted.

Errors

Most people will experience at least one diagnostic error in their lifetime, according to a 2015 report by the National Academies of Sciences, Engineering, and Medicine.

Causes and factors of error in diagnosis are:

- The manifestation of disease are not sufficiently noticeable
- A disease is omitted from consideration
- Too much significance is given to some aspect of the diagnosis
- The condition is a rare disease with symptoms suggestive of many other conditions
- The condition has a rare presentation

Lag Time

When making a medical diagnosis, a lag time is a delay in time until a step towards diagnosis of a disease or condition is made. Types of lag times are mainly:

- *Onset-to-medical encounter lag time*, the time from onset of symptoms until visiting a health care provider
- *Encounter-to-diagnosis lag time*, the time from first medical encounter to diagnosis

SOCIETY AND CULTURE

Social Context

Diagnosis can take many forms. It might be a matter of naming the disease, lesion, dysfunction or disability. It might be a management-naming or prognosis-naming exercise. It may indicate either degree of abnormality on a continuum or kind of abnormality in a classification. It’s influenced by non-medical factors such as power, ethics and financial incentives for patient or doctor. It can be a brief summation or an extensive formulation, even taking the form of a story or metaphor. It might be a means of communication such as a computer code through which it triggers payment, prescription, notification, information or advice. It might be pathogenic or salutogenic. It’s generally uncertain and provisional. Once a diagnostic opinion has been reached, the provider is able to propose a management plan, which will include treatment as well as plans for follow-up. From this point on, in addition to treating the patient’s condition, the provider can educate the patient about the etiology, progression, prognosis, other outcomes, and possible treatments of her or his ailments, as well as providing advice for maintaining health.

A treatment plan is proposed which may include therapy and follow-up consultations and tests to monitor the condition and the progress of the treatment, if needed, usually according to the medical guidelines provided by the medical field on the treatment of the particular illness. Relevant information should be added to the medical record of the patient. A failure to respond to treatments that would normally work may indicate a need for review of the diagnosis.

Concepts Related to Diagnosis

Sub-types of diagnoses include:

- *Clinical diagnosis:* A diagnosis made on the basis of medical signs and patient-reported symptoms, rather than diagnostic tests
- *Laboratory diagnosis:* A diagnosis based significantly on laboratory reports or test results, rather than the physical examination of the patient. For instance, a proper diagnosis of infectious diseases usually requires both an examination of signs and symptoms, as well as laboratory characteristics of the pathogen involved.
- *Radiology diagnosis:* A diagnosis based primarily on the results from medical imaging studies. Greenstick fractures are common radiological diagnoses.
- *Principal diagnosis:* The single medical diagnosis that is most relevant to the patient's chief complaint or need for treatment. Many patients have additional diagnoses.
- *Admitting diagnosis:* The diagnosis given as the reason why the patient was admitted to the hospital; it may differ from the actual problem or from the *discharge diagnoses*, which are the diagnoses recorded when the patient is discharged from the hospital.
- *Differential diagnosis:* A process of identifying all of the possible diagnoses that could be connected to the signs, symptoms, and lab findings, and then ruling out diagnoses until a final determination can be made.
- *Diagnostic criteria:* Designates the combination of signs, symptoms, and test results that the clinician uses to attempt to determine the correct diagnosis. They are standards, normally published by international committees, and they are designed to offer the best sensitivity and specificity possible, respect the presence of a condition, with the state-of-the-art technology.
- *Prenatal diagnosis:* Diagnosis work done before birth
- *Diagnosis of exclusion:* A medical condition whose presence cannot be established with complete confidence from history, examination or testing. Diagnosis is therefore by elimination of all other reasonable possibilities.
- *Dual diagnosis:* The diagnosis of two related, but separate, medical conditions or co-morbidities; the term almost always refers to a diagnosis of a serious mental illness and a substance addiction.
- *Self-diagnosis:* The diagnosis or identification of a medical conditions in oneself. Self-diagnosis is very common.
- *Remote diagnosis:* A type of telemedicine that diagnoses a patient without being physically in the same room as the patient.
- *Nursing diagnosis:* Rather than focusing on biological processes, a nursing diagnosis identifies people's responses to situations in their lives, such as a readiness to change or a willingness to accept assistance.
- *Computer-aided diagnosis:* Providing symptoms allows the computer to identify the problem and diagnose the user to the best of its ability. Health screening begins by identifying the part of the body where the symptoms are located; the computer cross-references a database for the corresponding disease and presents a diagnosis.
- *Overdiagnosis:* The diagnosis of "disease" that will never cause symptoms, distress, or death during a patient's lifetime
- *Wastebasket diagnosis:* A vague, or even completely fake, medical or psychiatric label given to the patient or to the medical records department for essentially non-medical reasons, such as to reassure the patient by providing an official-sounding label, to make the provider look effective, or to obtain approval for treatment. This term is also used as a derogatory label for disputed, poorly described,

overused, or questionably classified diagnoses, such as pouchitis and senility, or to dismiss diagnoses that amount to overmedicalization, such as the labeling of normal responses to physical hunger as reactive hypoglycemia.

- *Retrospective diagnosis:* The labeling of an illness in a historical figure or specific historical event using modern knowledge, methods and disease classifications.

HISTORY OF MEDICAL DIAGNOSIS

The history of medical diagnosis began in earnest from the days of Imhotep in ancient Egypt and Hippocrates in ancient Greece but is far from perfect despite the enormous bounty of information made available by medical research including the sequencing of the human genome. The practice of diagnosis continues to be dominated by theories set down in the early 20th century.

ANCIENT EGYPT

An Egyptian medical textbook, the Edwin Smith Papyrus written by Imhotep (fl. 2630-2611 BC), was the first to apply the method of diagnosis to the treatment of disease.

ANCIENT BABYLONIA

A Babylonian medical textbook, the *Diagnostic Handbook* written by Esagil-kin-apli (fl. 1069-1046 BC), introduced the use of empiricism, logic and rationality in the diagnosis of an illness or disease. The book made use of logical rules in combining observed symptoms on the body of a patient with its diagnosis and prognosis. He described the symptoms for many varieties of epilepsy and related ailments along with their diagnosis and prognosis which both played significant roles in the practice of Babylonian medicine.

ANCIENT CHINA

Predated by Babylonian and Egyptian medicine, traditional Chinese medicine (TCM) was described in an ancient Chinese text, the Yellow Emperor's Inner Canon or Huangdi Neijing which dates to the first or second century BCE.

The four diagnostic methods of TCM which are still being practiced today are inspection, listening and smelling, inquiry and palpation.

ANCIENT GREECE

Over two thousand years ago, Hippocrates recorded the association between disease and heredity. In similar fashion, Pythagoras noted the association between metabolism and heredity (allergy to Fava beans). The medical community, however, has only recently acknowledged the importance of genetics and its relevance to mainstream medicine.

ISLAMIC WORLD

The Arabic physician, Abu al-Qasim al-Zahrawi (Abulcasis), wrote on hematology in his *Al-Tasrif* (1000). He provided the first description on haemophilia, a hereditary genetic disorder, in which he wrote of an Andalusian family whose males died of bleeding after minor injuries. The Persian physician, Ibn Sina (Avicenna, 980-1037), in *The Canon of Medicine* (1025), pioneered the idea of a syndrome in the diagnosis of specific diseases.

MIDDLE AGES

Physicians used many different techniques to analyze the imbalance of the four humours in the body. Uroscopy was most widely used for diagnosing illness. Physicians would collect patients urine in a flask called “matula”. The matula was specific in shape and had four regions – circulus, superficies, substantia, and fundus – that corresponded to regions of the body. The circulus corresponded to the head; the superficies corresponded to the chest; the substantia corresponded to the abdomen; the fundus corresponded to the reproductive and urinary organs. Urine was inspected based on four criteria: colour, consistency, odour, and presence of precipitate.

Physicians analyzed the urine for the four criteria and used that to point out where there was an imbalance of the four humours based on the location in the matula. Physicians also examined blood via phlebotomy, they would observe the viscosity and colour of the blood as it was draining from the patient and/or contained in a vial. The colour and viscosity denoted whether the patient had an acute, major, or chronic disease; which also assisted the physician with the next course of action. Physicians would also observe a patients pulse by palpation; this technique was performed by carefully noting the rate, power, and tempo of a pulsing artery. By interpreting the pulse of the physician could diagnose the type of fever the patient had. Astrological diagnosis was the least used technique for diagnosing illness. Diagnosis was based on the position of the moon in relation to the constellations, which were associated with different regions of the body (head, arms, chest, etc.). Physicians would diagnose illness with the combined knowledge of zodiac signs and humoral medicine.

19TH CENTURY

In the first half of the 19th century, the well-known British physiologist Marshall Hall emphasized the necessity of maintaining a close relationship between the theory and practice in medicine. He wrote *On diagnosis* (1817) and *The Principles of Diagnosis* (1834).

THE OSLERIAN IDEAL

The ideals of William Osler, who transformed the practice of medicine in the early 1900s, were based on the principles of the diagnosis and treatment of disease. According to Osler, the functions of a physician were to be able to identify disease and its manifestations and to understand its mechanisms and how it may be prevented or cured. For his medical students he believed that the best textbook was the patient himself—analysis of morbid anatomy and pathology were the keys. The Oslerian ideal continues today as the basis of the doctor’s strategy is, “What disease does this patient have, and what is the best way for treatment?” The emphasis is on the classification of the disease in order to use the remedies available for its effects to be reversed or ameliorated. The human being in question is representative of a class of people with this type of disease; this person’s biological individuality is not given any great weight.

GARROD’S VIEW

The successor to William Osler as Regius Professor at Oxford was Archibald Garrod. Garrod echoed the observations of his Greek counterparts of two millennia ago, *...our chemical individualities are due to our chemical merits as well as our chemical shortcomings; and it is more nearly true to say that the factors which confer upon us our predispositions to and immunities from various mishaps which are spoken of as diseases, are inherent in our very chemical structure; and even in the molecular groupings which confer upon us our individualities, and which went into the making of the chromosomes from which we sprang.* Because Garrod practiced in the early 1900s, well before the knowledge of DNA encoding genes that in turn encoded proteins responsible for bodily structure and functions were discovered, it took some time before medicine could fully appreciate the fundamental importance of his concept of diagnosis.

PRESENT-DAY OSLERIAN PRACTICE

Whereas Osler laid the founding principles by which medicine should be practiced, Garrod placed these principles in a greater context of a chemical individuality that is inherited and is subject to the mechanisms of evolutionary selection. The Oslerian ideal of medical practice continues to dominate medical philosophy today. The patient is a collective of symptoms to be characterized and analyzed algorithmically in order to draw a diagnosis and subsequently produce a strategy of treatment. Medicine is about problems based solutions. In keeping with this philosophy, today's pathology reports provide a momentary snapshot of the patient's biochemical profile, highlighting the end result of the disease process.

INFLUENCE OF DNA TECHNOLOGY

Garrod's conception of biological individuality was confirmed with the advent of the sequencing of the human genome. Finally the subtle relationship between inheritance, individuality and environment became apparent via the variations detected in DNA. In each patient's DNA lies a script for how their bodies will change and become ill as well as how they will handle the assaults of the environment from the beginning of their life to its end. It is hoped that by knowing a patient's genes that the biological strengths and weaknesses in respect to these assaults will be revealed and disease processes can be predicted before they have the opportunity to manifest. Although knowledge in this area is far from complete, there are already medical interventions based on this. More importantly, the physician, forewarned with this knowledge can guide the patient towards appropriate lifestyle changes to anticipate and mitigate disease processes.

PHOTOACOUSTIC EFFECT

The photoacoustic effect or optoacoustic effect is the formation of sound waves following light absorption in a material sample. In order to obtain this effect the light intensity must vary, either periodically (*modulated light*) or as a single flash (*pulsed light*). The photoacoustic effect is quantified by measuring the formed sound (pressure changes) with appropriate detectors, such as microphones or piezoelectric sensors. The time variation of the electric output (current or voltage) from these detectors is the photoacoustic signal. These measurements are useful to determine certain properties of the studied sample. For example, in photoacoustic spectroscopy, the photoacoustic signal is used to obtain the actual absorption of light in either opaque or transparent objects. It is useful for substances in extremely low concentrations, because very strong pulses of light from a laser can be used to increase sensitivity and very narrow wavelengths can be used for specificity. Furthermore, photoacoustic measurements serve as a valuable research tool in the study of the heat evolved in photochemical reactions, particularly in the study of photosynthesis. Most generally, electromagnetic radiation of any kind can give rise to a photoacoustic effect. This includes the whole range of electromagnetic frequencies, from gamma radiation and X-rays to microwave and radio. Still, much of the reported research and applications, utilizing the photoacoustic effect, is concerned with the near ultraviolet/visible and infrared spectral regions.

OVERDIAGNOSIS

Overdiagnosis is the diagnosis of "disease" that will never cause symptoms or death during a patient's ordinarily expected lifetime. Overdiagnosis is a side effect of screening for early forms of disease. Although screening saves lives in some cases, in others it may turn people into patients unnecessarily and may lead to treatments that do no good and perhaps do harm. Given the tremendous variability that is normal in biology, it is inherent that the more one screens, the more incidental findings will be found. For a large percentage of them, the most appropriate medical response is to recognize them as something that does not require intervention; but determining which action a particular finding warrants ("ignoring", watchful waiting, or intervention) can be

very difficult, whether because the differential diagnosis is uncertain or because the risk ratio is uncertain (risks posed by intervention, namely, adverse events, versus risks posed by not intervening). Overdiagnosis occurs when a disease is diagnosed correctly, but the diagnosis is irrelevant. A correct diagnosis may be irrelevant because treatment for the disease is not available, not needed, or not wanted. Some people contend that the term “overdiagnosis” is inappropriate, and that “overtreatment” is more representative of the phenomenon. Because most people who are diagnosed are also treated, it is difficult to assess whether overdiagnosis has occurred in an individual. Overdiagnosis in an individual cannot be determined during life. Overdiagnosis is only certain when an individual remains untreated, never develops symptoms of the disease and dies of something else. Thus most of the inferences about overdiagnosis comes from the study of populations. Rapidly rising rates of testing and disease diagnosis in the setting of stable rates of the feared outcome of the disease (*e.g.* death) are highly suggestive of overdiagnosis. Most compelling, however, is evidence from a randomized trial of a screening test intended to detect pre-clinical disease. A persistent excess of detected disease in the tested group years after the trial is completed constitutes the best evidence that overdiagnosis has occurred. Although overdiagnosis is potentially applicable to the diagnosis of any disease, its origin is in cancer screening – the systematic evaluation of asymptomatic patients to detect early forms of cancer. The central harm of cancer screening is overdiagnosis – the detection of abnormalities that meet the pathologic definition of cancer (under the microscope) but will never progress to cause symptoms or death during a patient’s ordinarily expected lifetime.

NUCLEIC ACID TEST

A nucleic acid test (NAT) or nucleic acid amplification test (NAAT) is a technique utilized to detect a particular nucleic acid, virus, or bacteria which acts as a pathogen in blood, tissue, urine, etc. NAT system is a little different from other tests in that it detects genetic materials rather than antigens or antibodies. Detection of genetic materials allows an early diagnosis of a disease because the detection of antigens requires time for antigens to appear in the bloodstream. Since the amount of a certain genetic material is usually very small, NAT includes an amplification step of the genetic material. There are several ways of amplification including polymerase chain reaction (PCR), strand displacement assay (SDA), or transcription mediated assay (TMA). Virtually all nucleic acid amplification methods and detection technologies utilize the specificity of Watson-Crick base pairing; single-stranded probe or primer molecules capture DNA or RNA target molecules of complementary strands. Therefore, the design of probe strands is highly significant to raise the sensitivity and specificity of the detection. However, the mutants which forms the genetic basis for a variety of human diseases are usually slightly different from the normal nucleic acids. Often, they are only different in a single base, *e.g.*, insertions, deletions, and single-nucleotide polymorphisms (SNPs). In this case, imperfect probe-target binding can easily occur which results in false-positive outcomes. Lots of researches have been studied to achieve single-base specificity.

NEKREP CASE

The Nekrep Case was a court trial that followed the death of a 12-year-old boy, Bor Nekrep, due to a misdiagnosis in Slovenia in March 2008. The treating pediatrician misdiagnosed an inborn defect of the urea cycle (a very rare genetic condition). After the boy had been brought by his parents to the Paediatric Clinic in Maribor, the doctor, among other things, measured the level of ammonia in his blood and on the basis of a significantly increased level erroneously diagnosed Reye syndrome. Because of the misdiagnosis and a delay in appropriate treatment, the boy developed cerebral oedema and died eight days later.

ANTISACCADE TASK

Saccadic eye movement is primarily controlled by the frontal cortex. The anti-saccade (AS) task is a gross estimation of injury or dysfunction of the frontal lobe, by assessing the brain’s ability to inhibit the reflexive saccade.

PROCEDURE

To perform the anti-saccade task, a patient is asked to fixate on a motionless target (such as a small dot). A stimulus is then presented to one side of the target. The patient is asked to make a saccade in the direction away from the stimulus. For example, if a stimulus is presented to the left of the motionless target, the patient should look towards the right. Failure to inhibit a reflexive saccade is considered an error. Neurologic disorders affecting both the frontal cortex or the basal ganglia have shown impaired performance on the anti-saccade task. These include schizophrenia, Huntington's disease, Parkinson's disease, and progressive supranuclear palsy. Additionally, disseminated brain disease such as Alzheimer's disease or AIDS dementia also have an increased number of errors.

HISTORY

The anti-saccade test was initially described in 1978 by Peter Hallet when he was a faculty member at the Department of Physiology of the University of Toronto. Many other researchers have used this task, including Guitton et al. and Pierrot-Deseilligny et al. In Guitton's studies, the AS task was administered to patients whose dorsolateral prefrontal cortex was removed therapeutically for intractable epilepsy. This group was compared to healthy controls and patients whose temporal lobes had been removed, also for seizure relief. Results from Guitton's studies showed that only patients with frontal lobe lesions performed abnormally on the AS task. In contrast, studies by Pierrot-Deseilligny et al. correlated high error rates of AS to specific lesions in the dorsolateral prefrontal cortex (DLPFC). In this study, patients with unilateral infarctions of the posterior parietal lobe were compared against infarctions in 3 frontal regions: the supplementary motor area, frontal eye fields (FEF), and the DLPFC. Only patients with infarctions of the DLPFC showed statistically significant error rates.

CURRENT USE

Research in Pathophysiology

Saccadic eye movements and anti-saccadic eye movements are carried out by similar regions of the brain: the frontal eye field (FEF), the supplementary motor area (SMA), the thalamus and putamen. Anti-saccades require an intention to will the movement and inhibit a reflexive response. It was therefore thought that the dorsolateral prefrontal cortex played a greater role in anti-saccadic movements. This theory was disproved by O'Driscoll et al. in a study that used positron emission tomography to analyze brain activity during anti-saccadic movement.

The study showed that the DLPFC was equally activated in both saccadic and anti-saccadic movements. Anti-saccades required increased activation of the FEF, SMA and putamen. Currently, the AS task is utilized as a gross assessment of frontal lobe function in neurologic and psychiatric diseases. The task has a high sensitivity, however its specificity is low. Of note, children and adults over the age of 70 will physiologically have an increased rate of error on the anti-saccade task.

Psychological Research

Modified versions of the task, often termed as *emotional antisaccade task*, have been used in psychological and psychophysiological research to investigate the interaction between visual attention and the processing of emotions. Emotion drives attention: people are usually biased to direct eye movements towards emotional rather than neutral and dull stimuli. The emotional version of the antisaccade task uses emotional stimuli

(photographs or conditioned stimuli) as visual targets, requiring participants to look in the opposite direction. Performance in the task is a measure for attentional biases in healthy participants and various disorders such as mood disorders, addiction and social anxiety.

BODY FLUID

Body fluid, bodily fluids, or biofluids are liquids within the bodies of living people. In normal healthy men, the total body water is about 60 per cent of the total body weight; it is slightly lower in women. A 70 kg (160 pound) man, then, has 42 liters of water in his body. This is divided between ICF and ECF in a two-to-one ratio: 28 liters are inside cells and 14 liters are outside cells. This ECF compartment is divided into the fluid between cells - the interstitial fluid volume - and the vascular volume, also called the blood plasma volume. The vascular volume is divided into the venous volume and the arterial volume; and the arterial volume has a conceptually useful but unmeasurable subcompartment called the effective arterial blood volume. There are approximately 6 to 10 liters of lymph in the body, compared to 3.5 to 5 liters of blood.

LIST OF BODY FLUIDS

By type:

- Amniotic fluid
- Aqueous humour and vitreous humour
- Bile
- Blood
- Blood plasma
- Blood serum
- Breast milk
- Cerebrospinal fluid
- Cerumen (earwax)
- Chyle
- Chyme
- Endolymph and perilymph
- Exudates
- Feces -
- Female ejaculate
- Gastric acid
- Gastric juice
- Lymph
- Mucus (including *nasal drainage* and *phlegm*)
- Pericardial fluid
- Peritoneal fluid
- Pleural fluid
- Pus
- Rheum
- Saliva
- Sebum (skin oil)
- Serous fluid
- Semen
- Smegma

- Sputum
- Synovial fluid
- Sweat
- Tears
- Urine
- Vaginal secretion
- Vomit

By location:

- Intracellular fluid
- Extracellular fluid
 - (a) Intravascular fluid (blood plasma)
 - (b) Interstitial fluid
 - (c) Lymphatic fluid (sometimes included in interstitial fluid)
 - (d) Transcellular fluid

In plants:

- Plant exudates

HEALTH

Body fluid is the term most often used in medical and health contexts. Modern medical, public health, and personal hygiene practices treat body fluids as potentially unclean. This is because they can be vectors for infectious diseases, such as sexually transmitted diseases or blood-borne diseases. Universal precautions and safer sex practices try to avoid exchanges of body fluids. Body fluids can be analyzed in medical laboratory in order to find microbes, inflammation, cancers, etc.

Clinical Samples

Clinical samples are generally defined as non-infectious human or animal materials including blood, saliva, excreta, body tissue and tissue fluids, and also FDA-approved pharmaceuticals that are blood products. In medical contexts, it is a specimen taken for diagnostic examination or evaluation, and for identification of disease or condition.

Sampling

Methods of sampling of body fluids include:

- *Blood sampling* for any blood test, in turn including:
 - (a) Arterial blood sampling, such as radial artery puncture
 - (b) Venous blood sampling, also called venipuncture
- Lumbar puncture to sample cerebrospinal fluid
- Thoracocentesis to sample pleural fluid
- Amniocentesis to sample amniotic fluid

BODY FLUIDS IN ART

A relatively new trend in contemporary art is to use body fluids in art, though there have been rarer uses of blood (and perhaps feces) for quite some time, and Marcel Duchamp used semen decades ago. Examples include:

- The controversial *Piss Christ* (1987), by Andres Serrano, which is a photograph of a crucifix submerged in urine;

- In Janine Antoni's *Conduit* (2009) she created a copper cast gargoyle device that she could pee through on the top of the Chrysler Building, Antoni's urine acting as the patina.
- Andy Warhol's *Oxidations* series, begun in 1977, in which he invited friends to urinate onto a canvas of metallic copper pigments, so that the uric acid would oxidize into abstract patterns;
- *Self* (1991, recast 1996) by Marc Quinn, a frozen cast of the artist's head made entirely of his own blood;
- *Piss Flowers*, by Helen Chadwick (1991–92), are twelve white-enameled bronzes cast from cavities made by urinating in snow (though this might not be characterized as the use of bodily fluids *in* art, just their use in preparation);
- performances by Lennie Lee involving feces, blood, vomit from 1990
- many paintings by Chris Ofili, which make use of elephant dung (from 1992).
- Gilbert and George's *The Naked Shit Pictures* (1995)
- Hermann Nitsch and Das Orgien Mysterien Theatre use urine, feces, blood and more in their ritual performances.
- Franko B from 1990 blood letting performances.
- The cover of the Metallica's album *Load* is an original artwork entitled "Semen and Blood III", one of three photographic studies by Andres Serrano created in 1990 by mingling the artist's own semen and bovine blood between two sheets of Plexiglas.

CANARDS SIGN

Canards sign is a rare but positive sign during the physical examination for the diagnosis of acute splenomegaly, it is used to complement the more traditional Castell's sign. The patient often presents with generalised symptoms such as LUQ abdominal pain, fevers and rigour. Canard's sign is a specific sign for this diagnosis. Canard's sign can be completed by having the patient raise their arms above their head and bring them down to the side at the same time.

Canard's sign is a positive response when the patient experiences sudden pain in the left arm pit during this movement. Signs of splenomegaly may include a palpable left upper quadrant abdominal mass or splenic rub. It is more traditionally detected on physical examination by using Castell's sign or Traube's space, but an ultrasound can be used to confirm diagnosis. In patients where the likelihood of splenomegaly is high, the physical exam is not sufficiently sensitive to detect it; abdominal imaging is indicated in such patients.

CIT PROGRAMME TUMOR IDENTITY CARDS

The "Cartes d'Identité des Tumeurs (CIT)" programme, launched and financed by the French charity "Ligue Nationale contre le Cancer", aims at refining the molecular knowledge of multiple types of tumors with the prospect of improving or developing better targeted therapeutic approaches. The CIT programme mainly relies on the large-scale and systematic profiling of large cohorts of tumors at various molecular levels including at least the genome, the epigenome, and the transcriptome.

COMPANION DIAGNOSTIC

A companion diagnostic (CDx) is a diagnostic test used as a companion to a therapeutic drug to determine its applicability to a specific person. Companion diagnostics are co-developed with drugs to aid in selecting or excluding patient groups for treatment with that particular drug on the basis of their biological characteristics that determine responders and non-responders to the therapy. Companion diagnostics are developed based on companion biomarkers, biomarkers that prospectively help predict likely response or severe toxicity.

COORDINATED DIAGNOSTICS

Coordinated Diagnostics is the portion of the coordinated care/healthcare model which focuses on diagnostic workflow, real-time data flow, information systems, expertise, and informed decision making. When practiced, Coordinated Diagnostics integrates the diagnostic data and activities of care providers, testing facilities, information systems providers, diagnostic domain experts, payers, and patients. Coordinated Diagnostics maximizes the effective use of diagnostic information and resources across the health care continuum in order to improve patient care while reducing overall costs.

PRACTICE

When practicing Coordinated Diagnostics, actors within the health care continuum:

- Decide to perform particular diagnostic tests based on balancing the trade-offs between optimizing the patient's health outcome while minimizing the cost of the diagnostic testing, treatment episodes, and overall patient care across episodes.
- Interpret diagnostic results accurately and completely based on a complete set of patient information.
- Communicate diagnostic information in real time to health care continuum actors that can take action.
- Engage patients and patient advocates in the diagnostic testing, data exchange, decision making, and action taking activities.

In order to fully perform these activities, actors across the health care continuum need:

- A complete view of results and interpretations from prior diagnostic testing for the patient.
- Prior diagnostic tests ordered for the patient that have not yet been resulted.
- Relevant patient information such as medication history and family history.
- Informed methodologies and tools which use all available information to guide decision makers through Coordinated Diagnostic activities.
- Access to domain experts (e.g. pathologists) who provide consultative services.
- Feedback loops with organize and analyze historic diagnostic information so as to improve the efficacy of the Coordinated Diagnostic processes.

INDUSTRY APPLICABILITY

The concept of Coordinated Diagnostics as a component of Coordinated Care is new to the Health care industry and its applicability to various industry challenges such as EHR interoperability and the correlation of diagnostic results across disciplines makes it a useful organizing principle.

COSMED

COSMED is a privately held, international, medical technology corporation specialized in Metabolic and Cardio Pulmonary diagnostic testing equipment aimed for either professional or medical use in different application fields (Hospital, Sports Science, Health care, Professional sports and Wellness).

HISTORY

Founded in 1980 and based in Rome, Italy, COSMED started manufacturing portable spirometers and pulmonary function testing equipment. In the nineties the company introduced new systems for gas exchange measurements and patented the first portable system with telemetric data transmission for the measurement of metabolic parameters during field testing. During early 2000 COSMED started to expand international business incorporating subsidiaries and registered offices in several countries (USA, China, Germany, UK).

In 2011, COSMED USA, Inc. completed the acquisition of US company Life Measurement, Inc. (LMI), provider of body composition assessment equipment based on Air Displacement Plethysmography technology.

PRODUCTS

COSMED manufactures a full range of spirometers, Pulmonary Function Test (PFT), Cardio Pulmonary Exercise Test (CPET) and Indirect Calorimetry diagnostic equipment. Since 2011, with the acquisition of Life Measurement Inc. (LMI), COSMED manufactures also body composition assessment equipment for infants, children and adults. COSMED main brands include: BOD POD, PEA POD, K4b2, Fitmate, Quark, Pony FX, microQuark, Spiropalm, Omnia

APPEARANCES

COSMED products were selected to outfit the physiology assessment laboratory in the 23rd James Bond film *Skyfall*. COSMED products were used on-screen by the MI6 medical staff to assess James Bond's cardio respiratory performance and fitness level.

AWARDS

COSMED product Spiropalm 6MWT handheld spirometer won the 2013 Product of Outstanding Interest (POINT) Award from the European Respiratory Society (ERS) during the ERS Congress in Barcelona (Spain).

CRYOGLOBULINEMIA

Cryoglobulinemia, cryoglobulinaemia, or cryoglobulinemic disease, is a medical condition in which the blood contains large amounts of cryoglobulins – proteins (mostly immunoglobulins themselves) that become insoluble at reduced temperatures. This should be contrasted with cold agglutinins, which cause agglutination of red blood cells. Cryoglobulins typically precipitate at temperatures below normal body temperature – 37 degrees Celsius (99 degrees Fahrenheit) – and will dissolve again if the blood is heated. The precipitated clump can block blood vessels and cause toes and fingers to become gangrenous. While this disease is commonly referred to as cryoglobulinemia in the medical literature, it is better termed cryoglobulinemic disease for two reasons:

- Cryoglobulinemia is also used to indicate the circulation of (usually low levels of) cryoglobulins in the absence of any symptoms or disease and
- Healthy individuals can develop transient asymptomatic cryoglobulinemia following certain infections.

In contrast to these benign instances of circulating cryoglobulins, cryoglobulinemic disease involves the signs and symptoms of precipitating cryoglobulins and is commonly associated with various pre-malignant, malignant, infectious, or autoimmune diseases that are the underlying cause for production of the cryoglobulins.

CLASSIFICATION

Since the first description of cryoglobulinemia in association with the clinical triad of skin purpura, joint pain, and weakness by Meltzer et al in 1967, the percentage of cryoglobulinemic diseases described as essential cryoglobulinemia or idiopathic cryoglobulinemia, that is cryoglobulinemic disease that is unassociated with an underlying disorder, has fallen. Currently most cases of this disease are found to be associated with premalignant, malignant, infectious, or autoimmune disorders that are the known or presumed causes for the production of cryoglobulins. This form of non-essential or non-idiopathic cryoglobulinemic disease is classically grouped into three types according to the Brouet classification. The classification distinguishes three subtypes of cryoglobulinemic

diseases based on two factors, the class of immunoglobulins in the cryoglobulin and the association of the cryoglobulinemic disease with other disorder. The following table lists these three types of cryoglobulinemic disease, characterized on the monoclonal immunoglobulin(s) comprising the involved cryoglobulin, percentage of total cryoglobulinemic disease cases, and class of disorders associated for each type.

Type	Composition	Percent of cases	Association with other diseases
Type I	Monoclonal IgG, IgM, IgA, or their κ or λ light chains	10–15%	Hematological diseases, particularly MGUS, smoldering multiple myeloma, multiple myeloma, Waldenström's macroglobulinemia, and chronic lymphocytic leukemia
Type II	Monoclonal IgM plus polyclonal IgG or, rarely, IgA	50–60%	Infectious diseases, particularly hepatitis C infection, HIV infection, and Hepatitis C and HIV coinfection; hematological diseases particularly B cell disorders; autoimmune diseases
Type III	Polyclonal IgM plus polyclonal IgG or IgA	25–30%	Autoimmune diseases, particularly Sjögren's syndrome and less commonly systemic lupus erythematosus and rheumatoid arthritis; infectious diseases particularly HCV infection

The monoclonal or polyclonal IgM proteins involved in Types II and III cryoglobulinemic disease have rheumatoid factor activity. That is, they bind to polyclonal immunoglobulins, activate the blood complement system, and thereby form tissue deposits that contain IgM, IgG (or, rarely, IgA), and components of the complement system, including in particular complement component 4. The vascular deposition of these types of cryoglobulin-containing immune complexes and complement can cause a clinical syndrome of cutaneous small-vessel vasculitis characterized by systemic vasculitis and inflammation termed cryoglobulinemic vasculitis. Accordingly, type II and type III cryoglobulinemic diseases are often grouped together and referred to as mixed cryoglobulinemia or mixed cryoglobulinemic disease. The monoclonal IgM involved in Type I cryoglobulinemic diseases lacks rheumatoid factor activity. More recent high resolution protein electrophoresis methods have detected a small monoclonal immunoglobulin component in type III cryoglobulins and/or a micro-heterogeneous composition of oligo-clonal (*i.e.* more than one monoclonal) immunoglobulin components or immunoglobulins with structures that do not fit into any classifications in the cryoglobulins of ~10 per cent of type II and III disease cases. It has been proposed that these cases be termed an intermediate type II-III variant of cryoglobulinemic disease and that some of the type III cases associated with the expression of low levels of a one or more isotypes of circulating monoclonal immunoglobulin(s) are in transition to type II disease.

SIGNS AND SYMPTOMS

The clinical features of cryoglobulinemic disease can reflect those due not only to the circulation of cryoglobulins but also to any underlying hematological premalignant or malignant disorder, infectious disease, or autoimmune syndrome. The following sections of clinical features focuses on those attributed to the cryoglobulins. Cryoglobulins cause tissue damage by three mechanisms; they can:

- Increase blood viscosity thereby reducing blood flow to tissues to cause the hyperviscosity syndrome (*i.e.* headache, confusion, blurry or loss of vision, hearing loss, and epistaxis;
- Deposit in small arteries and capillaries thereby plugging these blood vessels and causing infarction and necrosis of tissues including in particular skin, distal extremities, and kidneys;
- In type II and type III disease, deposit on the epithelium of blood vessels and activate the blood complement system to form pro-inflammatory elements such as C5a thereby initiating the systemic vascular inflammatory reaction termed cryoglobulinemic vasculitis.

Essential cryoglobulinemic disease

The signs and symptoms in the increasingly rare cases of cryoglobulinemic disease that cannot be attributed to an underlying disease generally resemble those of patients suffering Type II and III (*i.e.* mixed) cryoglobulinemic disease.

Type I Cryoglobulinemic Disease

Signs and symptoms due to the cryoglobulins of type I disease reflect the hyperviscosity and deposition of cryoglobulins within the blood vessels which reduce or stop blood perfusion to tissues. These events occur particularly in cases where blood cryoglobulin levels of monoclonal IgM are high in patients with IgM MGUS, smoldering Waldenström's macroglobulinemia, or Waldenström's macroglobulinemia and in uncommon cases where the levels of monoclonal IgA, IgG, free κ light chains, or free λ light chains are extremely high in patients with non-IgM MGUS, non-IgM smoldering multiple myeloma, or multiple myeloma. The interruption of blood flow to neurological tissues can cause symptoms of confusion, headache, hearing loss, and peripheral neuropathy. Interruption of blood flow to other tissues in type I disease can cause cutaneous manifestations of purpura, blue discoloration of the arms or legs (acrocyanosis), necrosis, ulcers, and livedo reticularis; spontaneous nose bleeds, joint pain, membranoproliferative glomerulonephritis; and cardiovascular disturbances such as shortness of breath, inadequate levels of oxygen in the blood (hypoxemia), and congestive heart failure.

Types II and III Cryoglobulinemic Disease

Types II and III (or mixed or variant) cryoglobulinemic disease may also present with symptoms and signs of blood hyperviscosity syndrome and deposition of cryoglobulins within blood vessels but also include those attributable to cryoglobulinemic vasculitis. "Meltzer's triad" of palpable purpura, joint pain, and generalized weakness occurs in ~33 per cent of patients presenting with type II or type III disease. One or more skin lesions including palpable purpura, ulcers, digital gangrene, and areas of necrosis occur in 69-89 per cent of these mixed disease cases; less common findings include painful peripheral neuropathy (often manifesting as mononeuritis multiplex in 19-44 per cent of cases), kidney disease (primarily membranoproliferative glomerulonephritis (30 per cent), joint pain (28 per cent), and, less commonly, dry eye syndrome, Raynaud phenomenon (*i.e.* episodic painful reductions in blood flow to the fingers and toes). While the glomerulonephritis occurring in mixed disease appears to be due to inflammatory vasculitis, the glomerulonephritis occurring in type I disease appears due to the interruption of blood flow. The hematological, infectious, and autoimmune diseases underlying type II cryoglobulinemic disease and the infectious and autoimmune diseases underlying type III cryoglobulinemic disease are also critical parts of the disease's clinical findings.

PATHOPHYSIOLOGY

Cryoglobulins

Cryoglobulins consists of one or more of the following components: monoclonal or polyclonal IgM, IgG, IgA antibodies, monoclonal κ , or λ free light chain portions of these antibodies, and proteins of the blood complement system, particularly complement component 4 (C4). The particular components involved are a reflection of the disorders which are associated with, and considered to be the cause of, the cryoglobulinemic disease. The cryoglobulin compositions and disorder associations in cryoglobulinemic disease are as follows:

- Monoclonal IgM-based cryoglobulin occurs in cases of Waldenström's macroglobulinemia and the pre-malignant precursors to this cancer, IgM monoclonal gammopathy of undetermined significance and smoldering Waldenström's macroglobulinemia.
- Monoclonal IgG or, rarely, IgA, κ light chain, or λ light chain cryoglobulins occur in cases of multiple myeloma and the pre-malignant precursors to this cancer, non-IgM monoclonal gammopathy of undetermined significance and non-IgM smoldering multiple myeloma. Non-IgM monoclonal immunoglobulin-based cases of cryoglobulinemic disease are less commonly associated with other B-

cell lymphocytic diseases *viz.*, Non-Hodgkin lymphoma, Hodgkin's lymphoma, B-cell chronic lymphocytic leukemia, and Castleman's disease; they occur rarely in non-B cell hematological disorders such as myelodysplastic syndromes and chronic myelogenous leukemia. Among these purely monoclonal immunoglobulin causes of cryoglobulinemic disease, Waldenström macroglobulinemia and multiple myeloma together account for ~40 per cent of cases; their pre-malignant precursors account for ~44 per cent of cases; and the other cited hematological diseases account for ~16 per cent of cases.

- Mixtures of monoclonal or polyclonal IgM, IgG, and/or IgA along with blood complement proteins such as C4 are the cryoglobulins associated with cases of infectious diseases, particularly hepatitis C infection, HIV infection, and Hepatitis C and HIV coinfection, and, less commonly or rarely, with cases of other infectious diseases such as hepatitis B infection, hepatitis A infection, cytomegalovirus infection, Epstein–Barr virus infection, lyme disease, syphilis, lepromatous leprosy, Q fever, poststreptococcal nephritis, subacute bacterial endocarditis, coccidioidomycosis, malaria, schistosomiasis, echinococcosis, toxoplasmosis, and Kala-azar. These mixed-protein cryoglobulins are also associated with autoimmune diseases, particularly Sjogren's syndrome, less commonly systemic lupus erythematosus and rheumatoid arthritis, and rarely polyarteritis nodosa, systemic sclerosis, temporal arteritis, polymyositis, Henoch-Schonlein purpura, pemphigus vulgaris, sarcoidosis, inflammatory bowel diseases, and others. In these mixed-protein depositions, the monoclonal or polyclonal IgM typically possesses rheumatoid factor activity and therefore binds to the Fc region of polyclonal IgG antibodies, activates the blood complement system, and complexes with complement components to form precipitates composed of IgM, IgG or IgG, and complement components, particularly complement component 4 (C4).

DIAGNOSIS

Cryoglobulinemia and cryoglobulinemic disease must be distinguished from cryofibrinogenemia or cryofibrinogenemic disease, conditions which involve the cold-induced intravascular deposition of circulating native fibrinogens. The cryoglobulins in plasma or serum precipitate at lower temperatures (*e.g.* 4°C). Since cryofibrinogens are present in plasma but greatly depleted in serum, precipitation tests for them are positive in plasma but negative in serum. Cryofibrinogenemia is occasionally found in cases of cryoglobulinemic disease. Cryoglobulinemic disease must also be distinguished from frostbite as well as numerous other conditions that have a clinical (particularly cutaneous) presentation similar to cryoglobulinemic disease but are not exacerbated by cold temperature, *e.g.* dysfibrinogenemia and dysfibrinogenemic disease (conditions involving the intravascular deposition of genetically abnormal circulating fibrinogens), purpura fulminans, cholesterol emboli, warfarin necrosis, ecthyma gangrenosum, and various hypercoagulable states. Rheumatoid factor is a sensitive test for cryoglobulinemia. The precipitated cryoglobulins are examined by immunoelectrophoresis and immunofixation to detect and quantify the presence of monoclonal IgG, IgM, IgA, κ light chain, or λ light chain immunoglobins. Other routine tests include measuring blood levels of rheumatoid factor activity, complement C4, other complement components, and hepatitis C antigen. Biopsies of skin lesions and, where indicated, kidney or other tissues can help in determining the nature of the vascular disease (immunoglobulin deposition, cryoglobulinemic vasculitis, or, in cases showing the presence of cryofibrinogenemia, fibrinogen deposition). In all events, further studies to determine the presence of hematological, infections, and autoimmune disorders are conducted on the basis of these findings as well as each case's clinical findings.

TREATMENT

All patients with symptomatic cryoglobulinemia are advised to avoid, or protect their extremities, from exposure to cold temperatures. Refrigerators, freezers, and air-conditioning represent dangers of such exposure.

Asymptomatic Cryoglobulinemia

Individuals found to have circulating cryoglobulins but no signs or symptoms of cryoglobulinemic diseases should be evaluated for the possibility that their cryoglobulinemia is a transient response to a recent or resolving infection. Those with a history of recent infection that also have a spontaneous and full resolution of their cryoglobulinemia need no further treatment. Individuals without a history of infection and not showing resolution of their cryoglobulinemia need to be further evaluated. Their cryoglobulins should be analyzed for their composition of immunoglobulin type(s) and complement component(s) and examined for the presence of the premalignant and malignant diseases associated with Type I disease as well as the infectious and autoimmune diseases associated with type II and type III disease. A study conducted in Italy on >140 asymptomatic individuals found five cases of hepatitis C-related and one case of hepatitis b-related cryoglobulinemia indicating that a complete clinical examination of asymptomatic individuals with cryoglobulinemia offers a means for finding people with serious but potentially treatable and even curable diseases. Individuals who show no evidence of a disease underlying their cryoglobulinemia and who remain asymptomatic should be followed closely for any changes that may indicate development of cryoglobulinemic disease.

Severely Symptomatic Cryoglobulinemic Disease

People affected by the severest, often life-threatening, complications of cryoglobulinemic disease require urgent plasmapheresis and/or plasma exchange in order to rapidly reduce the circulating levels of their cryoglobulins. Complications commonly requiring this intervention include: hyperviscosity disease with severe symptoms of neurological (*e.g.* stroke, mental impairment, and myelitis) and/or cardiovascular (*e.g.*, congestive heart failure, myocardial infarction) disturbances; vasculitis-driven intestinal ischemia, intestinal perforation, cholecystitis, or pancreatitis, causing acute abdominal pain, general malaise, fever, and/or bloody bowel movements; vasculitis-driven pulmonary disturbances (*e.g.* coughing up blood, acute respiratory failure, X-ray evidence of diffuse pulmonary infiltrates caused by diffuse alveolar hemorrhage); and severe kidney dysfunction due to intravascular deposition of immunoglobulins or vasculitis. Along with this urgent treatment, severely symptomatic patients are commonly started on therapy to treat any underlying disease; this treatment is often supplemented with anti-inflammatory drugs such as corticosteroids (*e.g.*, dexamethasone) and/or immunosuppressive drugs. Cases where no underlying disease is known are also often treated with the latter corticosteroid and immunosuppressive medications.

Type I Cryoglobulinemic Disease

Treatment of Type I disease is generally directed towards treating the underlying pre-malignant or malignant disorder. This involves appropriate chemotherapy regimens which may include bortezomib (promotes cell death by apoptosis in cells accumulating immunoglobulins) in patients with monoclonal immunoglobulin-induced renal failure and rituximab (antibody directed against CD20 surface antigen-bearing lymphocytes) in patients with Waldenstrom's macroglobulinemia).

Type II and III Cryoglobulinemic Disease

Treatment of mixed cryoglobulinemic disease is, similar to type I disease, directed towards treating any underlying disorder. This includes malignant (particularly Waldenström's macroglobulinemia in type II disease), infectious, or autoimmune diseases in type II and III disease. Recently, evidence of hepatitis C infection has been reported in the majority of mixed disease cases with rates being 70-90 per cent in areas with high incidences of hepatitis C. The most effective therapy for hepatitis C-associated cryoglobulinemic disease consists of a

combination of anti-viral drugs, pegylated INFA and ribavirin; depletion of B cells using rituximab in combination with antiviral therapy or used alone in patients refractory to antiviral therapy has also proven successful in treating the hepatitis C-associated disease. Data on the treatment of infectious causes other than hepatitis C for the mixed disease are limited. A current recommendation treats the underlying disease with appropriate antiviral, anti-bacterial, or anti-fungal agents, if available; in cases refractory to an appropriate drug, the addition of immunosuppressive drugs to the therapeutic regimen may improve results.

Mixed cryoglobulinemic disease associated with autoimmune disorders is treated with immunosuppressive drugs: combination of a corticosteroid with either cyclophosphamide, azathioprine, or mycophenolate or combination of a corticosteroid with rituximab have been used successfully to treat mixed disease associated with autoimmune disorders.

DIAGNOSTIC OVERSHADOWING

Diagnostic overshadowing is the attribution of a person's symptoms to their mental condition, when such symptoms actually suggest a comorbid condition. The term was first used to describe the underdiagnosis of mental illness in people with intellectual disability. In recent years, the term has also been used when physical illnesses are overlooked in people with mental illness.

DIFFERENTIAL DIAGNOSIS

In medicine, a differential diagnosis is the distinguishing of a particular disease or condition from others that present similar clinical features. Differential diagnostic procedures are used by physicians and other trained medical professionals to diagnose the specific disease in a patient, or, at least, to eliminate any imminently life-threatening conditions. Often, each individual option of a possible disease is called a differential diagnosis (for example, acute bronchitis could be a differential diagnosis in the evaluation of a cough that ends up with a final diagnosis of common cold).

More generally, a differential diagnostic procedure is a systematic diagnostic method used to identify the presence of a disease entity where multiple alternatives are possible. This method is essentially a process of elimination or at least a process of obtaining information that shrinks the "probabilities" of candidate conditions to negligible levels, by using evidence such as symptoms, patient history, and medical knowledge to adjust epistemic confidences in the mind of the diagnostician (or, for computerized or computer-assisted diagnosis, the software of the system). Differential diagnosis can be regarded as implementing aspects of the hypothetico-deductive method, in the sense that the potential presence of candidate diseases or conditions can be viewed as hypotheses that physicians further determine as being true or false. Common abbreviations of the term "differential diagnosis" include DDx, ddx, DD, D/Dx, ΔΔ, or ΔΔχ.

GENERAL COMPONENTS

A standard of care differential diagnosis has four steps. Patient safety requires that the physician:

- Gathers all information about the patient and creates a symptoms list. The list can be in writing or in the physician's head, as long as they make a list.
- Lists all possible causes (*candidate conditions*) for the symptoms. Again, this can be in writing or in the physician's head but it must be done.
- Prioritizes the list by placing the most urgently dangerous possible causes at the top of the list.
- Rules out or treats possible causes, beginning with the most urgently dangerous condition and working down the list. *Rule out*—practically—means use tests and other scientific methods to determine that a candidate condition has a clinically negligible probability of being the cause.

If no diagnosis remains, it means either that the physician made an error, or that the condition is undocumented. The physician removes diagnoses from the list by observing and applying tests that produce different results, depending on which diagnosis is correct.

A mnemonic to help in considering multiple possible pathological processes is VINDICATE'M:

- Vascular
- Inflammatory/ Infectious
- Neoplastic
- Degenerative/ Deficiency/ Drugs
- Idiopathic/ Intoxication/ Iatrogenic
- Congenital
- Autoimmune/ Allergic/ Anatomic
- Traumatic
- Endocrine/ Environmental
- Metabolic

SPECIFIC METHODS

There are several methods for differential diagnostic procedures, and several variants among those. Furthermore, a differential diagnostic procedure can be used concomitantly or alternately with protocols, guidelines, or other diagnostic procedures (such as pattern-recognition or using medical algorithms). For example, in case of medical emergency, there may not be enough time to do any detailed calculations or estimations of different probabilities, in which case the ABC protocol (*Airway, Breathing and Circulation*) may be more appropriate. Later, when the situation is less acute, a more comprehensive differential diagnostic procedure may be adopted. The differential diagnostic procedure may be simplified if a “pathognomonic” sign or symptom is found (in which case it is almost certain that the target condition is present) or in the absence of a *sine qua non*- sign or symptom (in which case it is almost certain that the target condition is absent). A diagnostician can be selective, considering first those disorders that are more likely (a probabilistic approach), more serious if left undiagnosed and untreated (a prognostic approach), or more responsive to treatment if offered (a pragmatic approach). Since the subjective probability of the presence of a condition is never exactly 100 per cent or 0 per cent, the differential diagnostic procedure may aim at specifying these various probabilities to form indications for further action. The following are two methods of differential diagnosis, being based on epidemiology and likelihood ratios, respectively.

Epidemiology-based Method

One method of performing a differential diagnosis by epidemiology aims to estimate the probability of each candidate condition by comparing their probabilities to have occurred in the first place in the individual. It is based on probabilities related both to the presentation (such as pain) and probabilities of the various candidate conditions (such as diseases).

COVERAGE OF CANDIDATE CONDITIONS

The validity of both the initial estimation of probabilities by epidemiology and further workup by likelihood ratios are dependent of inclusion of candidate conditions that are responsible for as large part as possible of the probability of having developed the condition, and it is clinically important to include those where relatively fast initiation of therapy is most likely to result in greatest benefit. If an important candidate condition is missed, no method of differential diagnosis will supply the correct conclusion. The need to find more candidate

conditions for inclusion increases with increasing severity of the presentation itself. For example, if the only presentation is a deviating laboratory parameter and all common harmful underlying conditions have been ruled out, then it may be acceptable to stop finding more candidate conditions, but this would much more likely be unacceptable if the presentation would have been severe pain.

COMBINATIONS

If two conditions get high post-test probabilities, especially if the sum of the probabilities for conditions with known likelihood ratios become higher than 100 per cent, then the actual condition is a combination of the two. In such cases, that combined condition can be added to the list of candidate conditions, and the calculations should start over from the beginning.

To continue the example used above, let's say that the history and physical examination was indicative of cancer as well, with a likelihood ratio of 3, giving an Odds(PostH and E) of 0.057, corresponding to a P(PostH and E) of 5.4 per cent. This would correspond to a "Sum of known P(PostH and E)" of 101.5 per cent. This is an indication for considering a combination of primary hyperparathyroidism and cancer, such as, in this case, a parathyroid hormone-producing parathyroid carcinoma. A recalculation may therefore be needed, with the first two conditions being separated into "primary hyperparathyroidism without cancer", "cancer without primary hyperparathyroidism" as well as "combined primary hyperparathyroidism and cancer", and likelihood ratios being applied to each condition separately. In this case, however, tissue has already been resected, wherein a histopathologic examination can be performed that includes the possibility of parathyroid carcinoma in the examination (which may entail appropriate sample staining). Let's say that the histopathologic examination confirms primary hyperparathyroidism, but also showed a malignant pattern. By an initial method by epidemiology, the incidence of parathyroid carcinoma is estimated at about 1 in 6 million people per year, giving a very low probability before taking any tests into consideration. In comparison, the probability that a non-malignant primary hyperparathyroidism would have occurred at the same time as an unrelated non-carcinoma cancer that presents with malignant cells in the parathyroid gland is calculated by multiplying the probabilities of the two. The resultant probability is, however, much smaller than the 1 in 6 million. Therefore, the probability of parathyroid carcinoma may still be close to 100 per cent after histopathologic examination despite the low probability of occurring in the first place. Let's finally say that the diagnosis of parathyroid carcinoma resulted in an extended surgery that removed remaining malignant tissue before it had metastasized, and the patient lived happily ever after.

MACHINE DIFFERENTIAL DIAGNOSIS

Machine differential diagnosis is the use of computer software to partly or fully make a differential diagnosis. It may be regarded as an application of artificial intelligence. Many studies demonstrate improvement of quality of care and reduction of medical errors by using such decision support systems. Some of these systems are designed for a specific medical problem such as schizophrenia, Lyme disease or ventilator-associated pneumonia. Others such as ESAGIL, Iliad, QMR, DiagnosisPro, VisualDx, Isabel, ZeroMD, DxMate, Symptoma, and Physician Cognition are designed to cover all major clinical and diagnostic findings to assist physicians with faster and more accurate diagnosis. However, these tools all still require advanced medical skills to rate symptoms and choose additional tests to deduce the probabilities of different diagnoses.

HISTORY

The method of differential diagnosis was first suggested for use in the diagnosis of mental disorders by Emil Kraepelin. It is more systematic than the old-fashioned method of diagnosis by *gestalt* (impression).

ALTERNATIVE MEDICAL MEANINGS

'Differential diagnosis' is also used more loosely, to refer simply to a list of the most common causes of a given symptom, to a list of disorders similar to a given disorder, or to such lists when they are annotated with advice on how to narrow the list down (the book 'French's Index of Differential Diagnosis', ISBN 0-340-81047-5, is an example). Thus, a differential diagnosis in this sense is medical information specially organized to aid in diagnosis.

USAGE APART FROM IN MEDICINE

Methods similar to those of differential diagnostic processes in medicine are also used by biological taxonomists to identify and classify organisms, living and extinct. For example, after finding an unknown species, there can first be a listing of all potential species, followed by ruling out of one by one until, optimally, only one potential choice remains. Similar procedures may be used by plant and maintenance engineers, automotive mechanics, and used to be used in diagnosing faulty electronic circuitry.

DIRECTHIT

DirectHit is a pharmacodiagnostic test used to determine the tumor sensitivity or resistance to drug regimens recommended for the treatment of breast cancer by the National Comprehensive Cancer Network. It is a noninvasive test performed on small amounts of tissue removed during the original surgery lumpectomy, mastectomy, or core biopsy. DirectHit was developed by CCC Diagnostics Inc., a biotechnology company established by former researchers from Johns Hopkins University. DirectHit was launched on 14 January 2010. Currently, it is the only available test for predicting treatment outcomes for anticancer chemotherapy drugs for breast cancer.

TECHNOLOGY

DirectHit is based on quantitative immunofluorescence technology. According to recent studies, quantitative immunofluorescence (QIF) can be applied to the standardization of protein analysis, resulting in increased sensitivity and reproducibility. It is important to be able to quantitate the expression of predictive factors in breast cancer, because response to therapy is often dependent upon the concentration of particular proteins within the tissue. DirectHit utilizes monoclonal antibodies (mAb) corresponding to each biomarker to stain the tumor tissue samples. Slides are evaluated with a computerized, fluorescence microscopy system. Digital images are acquired through a CCD camera. An algorithm was developed to process and analyze the digital images allowing direct relation between the amount of a specific protein within the cancer cells of the tumor and response to the corresponding drug. Quantitative measurement of up to five signals in a single cell can be obtained.

CLINICAL TRIALS

The DirectHit Panel for Breast Cancer was validated in retrospective clinical trials conducted at The University of Maryland Greenebaum Cancer Center and Harborview Cancer Center in Baltimore, MD. One key biomarker was selected as a correspondent to each drug, These biomarkers were identified as playing a significant role in cancer processes and as bearing a mechanistic relation to the action of the specified drug. The drug groups used as treatment and their related biomarkers are as follows: Trastuzumab/HER-2/neu, Antiestrogens/Estrogen Receptor, Taxanes/beta-tubulin III, 5FU/Thymidylate Synthase. In these trials DirectHit analysis of estrogen receptor and HER-2/neu displayed a higher predictive accuracy for treatment outcomes with anti estrogen

drugs and Trastuzumab than standard methods. In addition DirectHit displayed exceptional predictive accuracy for chemotherapy response (88 per cent). DirectHit also displayed an extreme specificity for predictions of drug resistance (100 per cent).

COST

The cost of the DirectHit Panel for Breast Cancer is \$3500. The use of DirectHit could result in significant cost savings because physicians will be able to customize treatment regimens for patients, eliminating ineffective drugs that would have otherwise been prescribed. If a 50 per cent treatment failure rate is assumed for standard chemotherapy, then the use of the DirectHit Test Panel for Breast Cancer could result in potential savings of approximately \$466M in drug costs alone for 40,000 late stage cancer patients in the United States. If ancillary costs are also considered, DirectHit testing could result in a total savings of \$566M - \$666M.

DOT CANCELLATION TEST

The Dot cancellation test or Bourdon-Wiersma test is a commonly used test of combined visual perception and vigilance. The test has been used in the evaluation of stroke where subjects were instructed to cross out all groups of 4 dots on an A4 paper. The numbers of uncrossed groups of 4 dots, groups of dots other than 4 crossed, and the time spent (maximum, 15 minutes) were taken into account. The Group-Bourdon test, a modification of the Bourdon-Wiersma, is one of a number of psychometric tests which trainee train drivers in the UK are required to pass. The test is based on the work of French psychologist Benjamin B. Bourdon (1860–1943) and Dutch neurologist Enno Dirk Wiersma (nl) (1858–1940).

DREHMANN SIGN

The Drehmann sign describes a clinical test of examining orthopedic patients and is widely used in the functional check of the hip joint. It was first described by Gustav Drehmann (Breslau, 1869–1932). The Drehmann sign is positive if an unavoidable passive external rotation of the hip occurs when performing a hip flexion. In addition, an internal rotation of the respective hip joint is either not possible or accompanied by pain when forcefully induced. The positive Drehmann sign is a typical clinical feature in slipped capital femoral epiphysis (SCFE), the impingement syndrome, Legg–Calvé–Perthes disease or in osteoarthritis of the hip joint.

ELBOW EXTENSION TEST

The Elbow extension test is simple test that can be administered as part of the physical exam to help guide health care providers diagnosis and management of acute elbow fractures. The elbow extension test is performed when an elbow fracture, most commonly caused by trauma, is suspected as the source of pain and dysfunction.

ADMINISTERING THE TEST

The patient is asked to flex the shoulder to ninety degrees. Then fully the extend both arms while supinated. If the patient is unable to extend the dysfunctional arm then a further workup including radiography for acute fracture should be pursued.

UTILITY

The “Elbow Extension Test,” published in the *British Medical Journal* in December 2008, had a negative predictive value for fracture in 98 per cent of adults and 96 per cent of children in a study done in five separate Emergency Departments in England. In this setting the “Elbow Extension Test” proved useful in ruling out

fracture, reducing the number of radiographic tests and reducing the overall cost of care. Follow-up consultation is recommended in any case. Of course it is up to the clinician and patient's discretion to decide if more testing is warranted or if a three- to seven-day follow-up appointment or phone consultation to re-evaluate symptoms is sufficient.

Electrophysiological techniques for clinical diagnosis

Electrophysiological techniques for clinical diagnosis will discuss the techniques borrowed from electrophysiology used in the clinical diagnosis of subjects. There are many processes that occur in the body which produce electrical signals that can be detected. Depending on the location and the source of these signals, distinct methods and techniques have been developed to properly target them.

ROLE OF ELECTROPHYSIOLOGY IN CLINICAL MEDICINE

Electrophysiology has a very important role in ensuring accurate clinical diagnoses. Many neurological diseases cause symptoms that manifest far from the injured or diseased tissues. Locating and treating all affected areas of the body is essential for proper patient care. Electrophysiology allows for the investigation of abnormal electrical signals in the body's tissues. It provides quantitative data to clinicians, supporting diagnostic processes and evaluating treatment success. Often, biological measures such as electrophysiology are more useful in assessing symptom severity than existing clinical measurement scales. Their objective nature removes subjective assignment of scores to symptom severity, subsequently leading to better informed health care decisions.

ELECTROPHYSIOLOGICAL TECHNIQUES

There are various techniques available to study and measure the electromagnetic signals of the body. The brain, the heart and skeletal muscles are prime sources of electric and magnetic fields that can be recorded and the resulting patterns can give insight on what ailments the subject may have. These electrophysiological techniques are named according to what data is measured and sometimes the anatomical location of the sources as follows:

Electroencephalography (EEG)

Electroencephalography is the measurement of brain activity through the surface of the scalp. Electroencephalography data can be processed through analytical procedures and certain derived summary indices of these analyses are called quantitative electroencephalography (QEEG). Data from evoked potentials can also be used processed in certain ways that can be considered quantitative EEG as well. If QEEG data is mapped then it is a topographic QEEG (also known as brain electrical activity mapping or BEAM)

Electrocardiography (EKG)

The heart is the muscle that pumps oxygenated blood to the whole body. As a very active muscle, it has peculiar electrical activity that can be measured and analyzed. Electrocardiography is the measurement of these signals.

Electromyography (EMG)

Electromyography is the measurement and analysis of the electrical activity in skeletal muscles. This technique is useful for diagnosing the health of the muscle tissue and the nerves that control them. EMG measures action potentials, called Motor Unit Action Potentials (MUAPs), created during muscle contraction. A few common uses are determining whether a muscle is active or inactive during movement (onset of activity), assessing the velocity of nerve conduction, and the amount of force generated during movement. Of these uses, determining the onset of muscle activity has been shown to be the most accurate.

Event-related Potentials (ERPs)

The firing of neurons throughout the brain has been known to have localized relationships to certain functions, processes and reactions to stimuli. With proper equipment it is possible to locate where in the brain neurons have been activated and measure their event related potentials. Event-related potentials can be classified as either: sensory, motor or cognitive.

Evoked Potentials (EPs)

Measurement of spontaneous electrophysiological activity does not always provide the desired information from the signals of interest.

In such cases, the application of a stimulus to the desired target can produce transient evoked potentials that can provide further insight not obtained from solely passive recording methods such as EEG, ECG, EMG or MEG.

Magnetoencephalography (MEG)

The measurement of the naturally occurring magnetic fields produced by the brain's electrical activity is called magnetoencephalography. This method differs from magnetic resonance imaging in that it passively measures the magnetic fields without altering the body's magnetization. However, data from MEG and MRI can be combined to create images that approximately map the estimated location of the natural magnetic fields. This composite imaging process is called magnetic source imaging (MSI).

Nerve Conduction Study (NCS)

An NCS measures the electrical conduction velocity and other characteristics of nerves in the body.

DIAGNOSABLE CONDITIONS AND APPLICABLE METHODS

The toolset of available electrophysiological techniques has been carefully applied to the study of patients suffering from a wide variety of conditions in the hopes of finding novel and more reliable diagnosis. For some conditions the use of these methods in their diagnosis is standard, but for others their applicability for diagnosis is still in the research phase.

Some conditions where the usefulness of electrophysiological techniques has been determined or being studied with promising results follow:

Brainstem Lesions in Traumatic Brain Injury

In the event of a traumatic brain injury the presence of a brainstem lesion has a significant impact in the prognosis of the patient. Although the development of MRI has allowed for very effective detection of brainstem lesions, evoked potentials measurements are also an electrophysiological technique that has been used for over 30 years in this context.

Carpal Tunnel Syndrome (CTS)

The compression of the median nerve within the carpal canal of the wrist and the progression of symptoms resulting from this entrapment is known as carpal tunnel syndrome (CTS). Nerve conduction studies have been used as a control electrophysiological method in the development of better CTS diagnostic techniques.

Dementia

Dementia is a progressive, degenerative brain disease that impairs cognitive functions. Alzheimer's disease and other types of dementia diagnosis is being improved through the use of electroencephalogram (EEG) and event-related potentials(ERP).

Epilepsy

Abnormally excessive or synchronous neuronal activity in the brain can cause seizures. These symptoms are characteristic of the neurological disorder known as epilepsy. Epilepsy is typically diagnosed with an EEG test. However, the effectiveness of MEG in the diagnosis of neocortical epilepsy has also been established.

Essential Tremor

It is hard to diagnose essential tremor and differentiate it from other types of tremor. The burst discharge patterns of EMG signals is compared to the frequency and amplitude of videotaped tremors to evaluate and diagnose essential tremor.

Heart Disease

Heart disease is one of the leading causes of death in the world. Diagnosis of heart disease can require various non-electrophysiological methods since there are many possible ailments, but EKG can be used to detect some.

Spasticity (Cerebral Palsy and Stroke)

Spasticity is a velocity dependent resistance to stretch due to increases in gamma motor neuron activity. The most commonly affected muscles are those that oppose gravity, the elbow and wrist flexors, knee extensors and ankle plantarflexors. Spasticity is a side effect of multiple central nervous system disorders including Cerebral Palsy, Stroke, Multiple Sclerosis and spinal cord injuries and results in limited joint range of motion of the affected limb. Stroke is currently the leading cause of death in the United States. Ischemic stroke is the most common form, accounting for 85 per cent of all stroke cases. It involves restricted blood flow to areas of the brain, resulting from a blood clot or ruptured vessel, and ultimately causes brain tissue damage. Spasticity is a common side effect of this brain tissue damage and affects the lives of many stroke survivors. The increased muscle tone of spastic muscles encumbers goal-directed movements; impairing activities of daily living while additionally causing pain and discomfort. Valid, reliable, and sensitive measures of spasticity are necessary to accurately diagnose patients suffering from this neurophysiological condition and assess the effectiveness of various treatment modalities. Electromyography (EMG) has been proposed by multiple researchers as an alternative measurement technique to quantify spasticity. EMG is used to determine the onset of muscle activity in spastic muscles, allowing for comparison between affected patients and asymptomatic individuals. Measurement tools involving EMG are argued to be more sensitive than currently used clinical scales like the Modified Ashworth Scale (MAS) in detecting spasticity symptom severity. The use of EMG offers a quantitative value of severity as opposed to relying on subjective scoring protocols. Malhotra et al. (2008) used EMG and the MAS to determine which method was more effective in detecting spasticity of the wrist flexors. In their sample of 100 patients (median age 74 years), they found that using EMG to detect onset of muscle activity during multiple passive stretch repetitions of the wrist at various velocities, successfully detected spasticity in 87 patients while the MAS only detected spasticity in 44 of these individuals. These findings support the use of EMG as a more sensitive diagnostic tool than the MAS and therefore advantageous when used in a clinical

setting. Similarly, EMG was successful in detecting elbow flexor spasticity symptom improvement in six post-stroke subjects (mean age 54.16 ± 7.9 years) following a neural mobilization technique of the median nerve. At 90 degrees of flexion and full extension of the elbow, muscle activity decreased from 17 per cent pre-treatment to 11 per cent post-treatment. These patients had initial MAS scores of 1 or 2, and the MAS was incapable of detecting the same symptom improvements post-treatment as seen with the EMG method. Cerebral palsy is another group of disorders caused by abnormal brain development or damage to the developing brain, which may result in spasticity. EMG has also been successfully used to quantify spasticity in this patient population. Using surface EMG and a torque motor, Levin et al. (1994, 2000) demonstrated a direct relationship between the velocity of stretch and the onset of spasticity. They incorporated EMG in their use of a novel motorized manipulandum due to its ability to detect stretch reflex activity and thus incorporate the velocity-dependent nature of spasticity in the resulting diagnostic values.

Multiple Sclerosis

The demyelination and scarring of axons in the neurons of the nervous system can affect their conduction properties and seriously harm the normal communication of the brain with the rest of the body. Multiple sclerosis (MS) is a disease that causes this deterioration of the myelin sheath. There isn't a unique test to diagnose MS and several studies must be combined to determine the presence of this disease. However, visual evoked potentials do play a role in the whole diagnostic process.

Parkinson's Disease

Parkinson's disease is a degenerative ailment that affects the central nervous system and is typically identified initially by its motor related symptoms. Accurate differentiation of PD from any other neurological disorder and the identification of the disease course is important in establishing an appropriate antiparkinsonian therapy. In the diagnostic role, surface EMG is a very informative method used to obtain relevant quantitative characteristics.

EXTERNAL CAUSE

In medicine, an external cause is a reason for the existence of a medical condition which can be associated with a specific object or acute process that was caused by something outside the body. They are classified as "E codes" in ICD 9. External Cause of Injury Codes (E codes) are ICD-9-CM codes or ICD-10 codes that are used to define the mechanism of death or injury, along with place of occurrence of the event. E codes are assigned on death certificates based on the manner of death. ICD-10 codes in the range V01–X59 refer to unintentional injuries. Codes in the range X60–X84 refer to intentional self-harm. Codes in the range Y85–Y09 refer to assault, and codes in the range Y10–Y34 refer to events of undetermined intent. E codes are well collected on death certificate data, but less so on hospital discharge data. Numerous initiatives have increased the percentage of records coded (CDC, MMWR March 28, 2008/ Vol. 57/ No. RR-1).

FAST (STROKE)

FAST is an acronym used as a mnemonic to help detect and enhance responsiveness to the needs of a person having a stroke. The acronym stands for Facial drooping, Arm weakness, Speech difficulties and Time to call emergency services.

- *Facial drooping:* A section of the face, usually only on one side, that is drooping and hard to move. This can be recognized by a crooked smile.
- *Arm weakness:* The inability to raise one's arm fully

- *Speech difficulties:* An inability or difficulty to understand or produce speech
- *Time:* If any of the symptoms above are showing, time is of the essence; call the emergency services or go to the hospital.

HISTORY

The FAST was developed in the UK in 1998 by a group of stroke physicians, ambulance personnel, and an emergency department physician and was designed to be an integral part of a training package for ambulance staff. The FAST was created to expedite administration of intravenous tissue plasminogen activator to patients within 3 hours of acute stroke symptom onset. The instruments at this time with most evidence of validity were the Cincinnati Prehospital Stroke Scale (CPSS) and the Los Angeles Prehospital Stroke Screen (LAPSS).

FINE-NEEDLE ASPIRATION

Fine-needle aspiration (FNA) is a diagnostic procedure used to investigate lumps or masses. In this technique, a thin (23-25 gauge), hollow needle is inserted into the mass for sampling of cells that, after being stained, will be examined under a microscope (biopsy). The sampling and biopsy considered together are called fine-needle aspiration biopsy (FNAB) or fine-needle aspiration cytology (FNAC) (the latter to emphasize that any aspiration biopsy involves cytopathology not histopathology). Fine-needle aspiration biopsies are very safe, minor surgical procedures. Often, a major surgical (excisional or open) biopsy can be avoided by performing a needle aspiration biopsy instead. In 1981, the first fine-needle aspiration biopsy in the United States was done at Maimonides Medical Center, eliminating the need for surgery and hospitalization. Today, this procedure is widely used in the diagnosis of cancer and inflammatory conditions. A needle aspiration biopsy is safer and less traumatic than an open surgical biopsy, and significant complications are usually rare, depending on the body site. Common complications include bruising and soreness. There is a risk, because the biopsy is very small (only a few cells), that the problematic cells will be missed, resulting in a false negative result. There is also a risk that the cells taken will not enable a definitive diagnosis.

MEDICATION USES

This type of sampling is performed for one of two reasons:

- A biopsy is performed on a lump or a tissue mass when its nature is in question.
- For known tumors, this biopsy is performed to assess the effect of treatment or to obtain tissue for special studies.

When the lump can be felt, the biopsy is usually performed by a cytopathologist or a surgeon. In this case, the procedure is usually short and simple. Otherwise, it may be performed by an interventional radiologist, a doctor with training in performing such biopsies under x-ray or ultrasound guidance. In this case, the procedure may require more extensive preparation and take more time to perform. Also, fine-needle aspiration is the main method used for chorionic villus sampling, as well as for many types of body fluid sampling. It is also used for ultrasound-guided aspiration of breast abscess, of breast cysts, and of seromas.

PREPARATION

The skin above the area to be biopsied is swabbed with an antiseptic solution and draped with sterile surgical towels. The skin, underlying fat, and muscle may be numbed with a local anesthetic, although this is often not necessary with superficial masses. After locating the mass for biopsy, using x-rays or palpation, a special needle of very fine diameter is passed into the mass. The needle may be inserted and withdrawn several times. There are many reasons for this:

- One needle may be used as a guide, with the other needles placed along it to achieve a more precise position.
- Sometimes, several passes may be needed to obtain enough cells for the intricate tests which the cytopathologists perform.

After the needles are placed into the mass, cells are withdrawn by aspiration with a syringe and spread on a glass slide. The patient's vital signs are taken again, and the patient is removed to an observation area for about 3 to 5 hours.

- For biopsies in the breast, ultrasound-guided fine needle biopsy is the most common. The biopsy is advised.

POST-OPERATIVE CARE AND COMPLICATIONS

As with any surgical procedure, complications are possible, but major complications due to thin needle aspiration biopsies are fairly uncommon, and when complications do occur, they are generally mild. The kind and severity of complications depend on the organs from which a biopsy is taken or the organs gone through to obtain cells. After the procedure, mild analgesics are used to control post-operative pain. Aspirin or aspirin substitutes should not be taken for 48 hours after the procedure (unless aspirin is prescribed for a cardiac or neurological condition). Since sterility is maintained throughout the procedure, infection is rare. But should an infection occur, it will be treated with antibiotics. Bleeding is the most common complication of this procedure. A slight bruise may also appear. If a lung or kidney biopsy has been performed, it is very common to see a small amount of blood in sputum or urine after the procedure. Only a small amount of bleeding should occur. During the observation period after the procedure, bleeding should decrease over time. If more bleeding occurs, this will be monitored until it subsides. Rarely, major surgery will be necessary to stop the bleeding.

Other complications depend upon the body part on which the biopsy takes place:

- Lung biopsies are frequently complicated by pneumothorax (collapsed lung). This complication can also accompany biopsies in the upper abdomen near the base of the lung. About one-quarter to one-half of patients having lung biopsies will develop pneumothorax. Usually, the degree of collapse is small and resolves on its own without treatment. A small percentage of patients will develop a pneumothorax serious enough to require hospitalization and placement of a chest tube for treatment. Although it is impossible to predict in whom this will occur, collapsed lungs are more frequent and more serious in patients with severe emphysema and in patients in whom the biopsy is difficult to perform.
- For biopsies of the liver, bile leakages may occur, but these are quite rare.
- Pancreatitis (inflammation of the pancreas) may occur after biopsies in the area around the pancreas.
- In biopsies in the area of the breast, bleeding and bruising may occur, less frequently also infection (rarely) or (very rarely, and only if performed near the chest wall) pneumothorax.
- Deaths have been reported from needle aspiration biopsies, but such outcomes are extremely rare.

CRITICISM

A recent study showed that in one case a needle biopsy of a liver tumor resulted in spread of the cancer along the path of the needle, and concluded that needle aspiration was dangerous and unnecessary. The conclusions drawn from this paper were strongly criticized subsequently.

FLOW-MEDIATED DILATION

Flow-mediated dilation (FMD) refers to dilation (widening) of an artery when blood flow increases in that artery. The primary cause of FMD is release of nitric oxide by endothelial cells. To determine FMD, brachial

artery dilation following a transient period of forearm ischemia is measured using ultrasound. Because the value of FMD can be compromised when improperly applied, attempts have been made to standardize the methodology for measuring FMD.

CLINICAL SIGNIFICANCE

FMD is a noninvasive measure of blood vessel health (endothelial dysfunction) which is at least as predictive of cardiovascular disease as traditional risk factors. Improved FMD results in greater perfusion and oxygen supply to peripheral tissue. An Israeli study of 618 healthy found FMD an excellent predictor of long-term adverse cardiovascular events. Participants with below-mean FMD were 278 per cent more likely to experience cardiovascular events during the 4.6 year average follow-up period than participant with below-mean FMD (95 per cent Confidence Interval: 135-571 per cent, p -value<0.001).

EFFECTS OF EXERCISE

A study of healthy young men who normally take over 10,000 steps per day, but were restricted to less than 5,000 steps per day for five days showed impaired FMD in the popliteal (leg) artery, but not the brachial (arm) artery. The reduction of leg FMD caused by prolonged sitting can be reduced by fidgeting (periodic leg movement). An eight-week programme of brisk walking resulted in a 50 per cent increase in brachial artery FMD in middle-aged and older men, but failed to produce this benefit in estrogen-deficient post-menopausal women. Forty-five minutes of cycling exercise before sitting has been shown to eliminate the impaired leg FMD due to three hours of sitting. Athletes over age 40 show greater FMD than their age-matched peers. A meta-analysis of 182 subjects showed twice the improvement in FMD resulting from high-intensity interval training compared to endurance training.

FRAX

FRAX is a diagnostic tool used to evaluate the 10-year probability of bone fracture risk. It was developed by the University of Sheffield. FRAX integrates clinical risk factors and bone mineral density at the femoral neck to calculate the 10-year probability of hip fracture and the 10-year probability of a major osteoporotic fracture (clinical spine, forearm, hip or shoulder fracture). The models used to develop the FRAX diagnostic tool were derived from studying patient populations in North America, Europe, Latin America, Asia and Australia.

COMPONENTS

The parameters included in a FRAX assessment are:

- Country
- Age
- Sex
- Weight
- Height
- Previous fracture
- Hip fracture in the subject's mother or father
- Smoking
- Glucocorticoid treatment
- Rheumatoid arthritis
- Disease strongly associated with osteoporosis

- Alcohol intake of 3 or more standard drinks per day
- Bone mineral density (BMD) of the femoral neck

AVAILABILITY AND USAGE

FRAX is freely accessible online, and commercially available as a desktop application, in paper-form as a FRAX Pad, as an iPhone application, and as an Android application. The tool is compatible with 58 models for 53 countries, and is available in 28 languages. FRAX is incorporated into many national guidelines around the world, including those of Belgium, Canada, Japan, Netherlands, Poland, Sweden, Switzerland, UK (NOGG), and US (NOF). FRAX assessments are intended to provide guidance for determining access to treatment in health care systems.

GENDER-BIAS IN MEDICAL DIAGNOSIS

Gender-biased diagnosing is a controversial sociological observation which asserts that medical and psychological diagnosis may be influenced by the gender of the patient. Several studies have found evidence of differential diagnosis for patients with similar ailments but of different sexes.

MEDICAL DIAGNOSIS

The phenomenon may affect physical diagnosis. Women are more likely to be given a diagnosis of psychosomatic nature for a physical ailment than men, despite presenting with similar symptoms. Women sometimes have trouble being taken seriously by physicians when suffering from medically unexplained illness, and report difficulty receiving appropriate medical care for their illnesses because doctors repeatedly diagnose their physical complaints as related to psychiatric problems. Clinical offices that rely on health care routines become less distinct due to biased medical knowledge of gender. There is a distinct differentiation between gender and sex in the medical sense. Gender is the societal construction of what femininity and masculinity is, whereas, sex is the biological aspect that defines the dichotomy of female and male. The way of lifestyle and the place in society are often considered when diagnosing patients. According to traditional medical studies, most of these medical studies were done on men thus overlooking many issues that were related to women's health. This topic alone sparked controversy and brought about question to the medical standard of our time. Research that was done on diseases that affected women more were less funded than those diseases that affected men and women equally.

There was an example of gender bias in the psychiatric field as well, Hamburg notes that, "psychiatrists would diagnose women with depression and then, eventually psychiatrists would begin to assume that women were more depressed than men due to the fact that the patients that were examined by the psychiatrists were women and they had similar symptoms. As for the men, they were diagnosed with drug or alcohol problems and they were thrown out of the study." As for the Women's Health Equity Act that was passed in 1993, this Act gave women the chance to participate in medical studies and examine the gender differences. Before the Women's Health Equity Act was introduced there was no research done on infertility, breast cancer, and ovarian cancer which were essential to women's health.

PSYCHOLOGICAL DIAGNOSIS

There is a suggestion that assumptions regarding gender specific behavioural characteristics can lead to a diagnostic system which is biased. The issue of gender bias with regard to *Diagnostic and Statistical Manual of Mental Disorders* (DSM) personality disorder criteria has been controversial and widely debated. The fourth DSM (4th ed., text revision; DSM-IV-TR; American Psychiatric Association, 2000) makes no explicit statement

regarding gender bias among the ten personality disorders (PDs), but it does state that six PDs (antisocial, narcissistic, obsessive-compulsive, paranoid, schizotypal, schizoid) are more frequently found in men. Three others (borderline, histrionic, dependent) are more frequent in women. Avoidant is equally common in men and women. There are many ways to interpret differential prevalence rates as a function of gender. Some critics have argued that they are an artifact of gender bias. In other words, the PD criteria assume unfairly that stereotypical female characteristics are pathological. The results of this study conclude with no indication of gender-biased criteria in the borderline, histrionic, and dependent PDs. This is in contrast with what is predicted by critics of these disorders, who suggest they are biased against women. It is possible, however, that other sources of bias, including assessment and clinical bias, are still at work in relation to these disorders. The results do show that the group means are higher in women than in men, an expected result considering the higher prevalence rate of these disorders for women.

The original purpose of the DSM-IV was to provide an accurate classification of psychopathology, not to develop a diagnostic system that will, democratically, diagnose as many men with a personality disorder as women". However, if the criteria are to serve equally as indicators of disorder for both men and women, it will be important to establish that the implications of these criteria for functional impairment are comparable for both sexes. Whereas it is plausible that there are gender-specific expressions of these disorders, DSM-IV criteria that function differentially for men and women can systematically over-pathologize or under-represent mental illness in a particular gender. The present study is limited by the investigation of only four personality disorders and the lack of inclusion of additional diagnoses that have also been controversial in the gender bias debate (such as dependent and histrionic personality disorders), although it offers a clearly articulated methodology for studying this possibility. In addition, it provides an examination of a clinical sample of substantial size and uses functional assessments that cut across multiple functional domains and multiple assessment methods. Our results indicate that BPD criteria showed some evidence of differential functioning between genders on global functioning, although there is little evidence of sex bias within the diagnostic criteria for avoidant, schizotypal, or obsessive-compulsive personality disorders. Further investigation and validation across sexes for those disorders would be an important direction of future research.

Considerable evidence indicates a prominent role for trauma-related cognitions in the development and maintenance of posttraumatic stress disorder (PTSD) symptoms. The present study utilized regression analysis to examine the unique relationships between various trauma-related cognitions and PTSD symptoms after controlling for gender and measures of general affective distress in a large sample of trauma-exposed college students. In terms of trauma-related cognitions, only negative cognitions about the self were related to PTSD symptom severity. Gender and anxiety symptoms were also related to PTSD symptom severity. Theoretical implications of the results are consistent with previous studies on the relationship between PTSD and negative cognitions, the self, world, and blame subscales of the PTCI were significantly related to PTSD symptoms. The study correlations indicated that increased negative trauma-related cognitions were related to more severe PTSD symptoms. Also consistent with previous reports, correlations also indicated that gender was related to PTSD symptom severity, such that women had more severe PTSD symptoms. PTSD symptom severity was also positively related to depression, anxiety, and stress reactivity.

Distinguishing between borderline personality disorder (BPD) and posttraumatic stress disorder (PTSD) is often challenging, especially when the client has experienced a trauma such as childhood sexual abuse (CSA), which is strongly linked to both disorders. Although the individual diagnostic criteria for these two disorders do not overlap substantially, patients with either of these disorders can display similar clinical pictures. Both patients with BPD and PTSD may present as aggressive towards self or others, irritable, unable to tolerate emotional extremes, dysphoric, feeling empty or dead, and highly reactive to mild stressors. Despite having similar clinical pictures, PTSD and BPD are regarded differently by many clinicians. Results from a 2009 study concluded that patient gender does not affect diagnosis. This finding is consistent with research suggesting that

women are not more likely to be given the BPD diagnosis, all else being equal, though it contradicts other findings from studies that have used similar case vignettes. Nor did the data support an effect of clinician gender or age on diagnosis.

A 2012 study examined gender-specific associations between trauma cognitions, alcohol cravings and alcohol-related consequences in individuals with dually diagnosed PTSD and alcohol dependence (AD). Participants had entered a treatment study for concurrent PTSD and AD; baseline information was collected from participants about PTSD-related cognitions in three areas:

- Negative Cognitions About Self,
- Negative Cognitions About the World, and
- Self-Blame. Information was also collected on two aspects of AD: alcohol cravings and consequences of AD. Gender differences were examined while controlling for PTSD severity. The results indicate that Negative Cognitions About Self are significantly related to alcohol cravings in men but not women, and that interpersonal consequences of AD are significantly related to Self-Blame in women but not in men. These findings suggest that for individuals with comorbid PTSD and AD, psychotherapeutic interventions that focus on reducing trauma-related cognitions are likely to reduce alcohol cravings in men and relational problems in women.

IMMUNOLABELING

Immunolabeling is a biochemical process that enables the detection and localization of an antigen to a particular site within a cell, tissue, or organ. Antigens are organic molecules, usually proteins, capable of binding to an antibody. These antigens can be visualized using a combination of antigen-specific antibody as well as a means of detection, called a tag, that is covalently linked to the antibody. If the immunolabeling process is meant to reveal information about a cell or its substructures, the process is called immunocytochemistry. Immunolabeling of larger structures is called immunohistochemistry. There are two complex steps in the manufacture of antibody for immunolabeling. The first is producing the antibody that binds specifically to the antigen of interest and the second is fusing the tag to the antibody. Since it is impractical to fuse a tag to every conceivable antigen-specific antibody, most immunolabeling processes use an indirect method of detection. This indirect method employs a primary antibody that is antigen-specific and a secondary antibody fused to a tag that specifically binds the primary antibody. This indirect approach permits mass production of secondary antibody that can be bought off the shelf. Pursuant to this indirect method, the primary antibody is added to the test system. The primary antibody seeks out and binds to the target antigen. The tagged secondary antibody, designed to attach exclusively to the primary antibody, is subsequently added. Typical tags include: a fluorescent compound, gold beads, a particular epitope tag, or an enzyme that produces a colored compound. The association of the tags to the target via the antibodies provides for the identification and visualization of the antigen of interest in its native location in the tissue, such as the cell membrane, cytoplasm, or nuclear membrane. Under certain conditions the method can be adapted to provide quantitative information. Immunolabeling can be used in pharmacology, molecular biology, biochemistry and any other field where it is important to know of the precise location of an antibody-bindable molecule.

INDIRECT VS. DIRECT METHOD

There are two methods involved in immunolabeling, the direct and the indirect methods. In the direct method of immunolabeling, the primary antibody is conjugated directly to the tag. The direct method is useful in minimizing cross-reaction, a measure of nonspecificity that is inherent in all antibodies and that is multiplied with each additional antibody used to detect an antigen. However, the direct method is far less practical than the indirect method, and is not commonly used in laboratories, since the primary antibodies must be covalently

labeled, which require an abundant supply of purified antibody. Also, the direct method is potentially far less sensitive than the indirect method. Since several secondary antibodies are capable of binding to different parts, or domains, of a single primary antibody binding the target antigen, there is more tagged antibody associated with each antigen. More tag per antigen results in more signal per antigen. Different indirect methods can be employed to achieve high degrees of specificity and sensitivity. First, two-step protocols are often used to avoid the cross-reaction between the immunolabeling of multiple primary and secondary antibody mixtures, where secondary fragment antigen-binding antibodies are frequently used. Secondly, haptenylated primary antibodies can be used, where the secondary antibody can recognize the associated hapten. The hapten is covalently linked to the primary antibody by succinyl imide esters or conjugated IgG Fc-specific Fab sections. Lastly, primary monoclonal antibodies that have different Ig isotypes can be detected by specific secondary antibodies that are against the isotype of interest.

ANTIBODY BINDING AND SPECIFICITY

Overall, antibodies must bind to the antigens with a high specificity and affinity. The specificity of the binding refers to an antibody's capacity to bind and only bind a single target antigen. Scientists commonly use monoclonal antibodies and polyclonal antibodies, which are composed of synthetic peptides. During the manufacture of these antibodies, antigen specific antibodies are sequestered by attaching the antigenic peptide to an affinity column and allowing nonspecific antibody to simply pass through the column. This decreases the likelihood that the antibodies will bind to an unwanted epitope of the antigen not found on the initial peptide. Hence, the specificity of the antibody is established by the specific reaction with the protein or peptide that is used for immunization by specific methods, such as immunoblotting or immunoprecipitation.

In establishing the specificity of antibodies, the key factor is the type of synthetic peptides or purified proteins being used. The lesser the specificity of the antibody, the greater the chance of visualizing something other than the target antigen. In the case of synthetic peptides, the advantage is the amino acid sequence is easily accessible, but the peptides do not always resemble the 3-D structure or post-translational modification found in the native form of the protein. Therefore, antibodies that are produced to work against a synthetic peptide may have problems with the native 3-D protein. These types of antibodies would lead to poor results in immunoprecipitation or immunohistochemistry experiments, yet the antibodies may be capable of binding to the denatured form of the protein during an immunoblotting run. On the contrary, if the antibody works well for purified proteins in their native form and not denatured, an immunoblot cannot be used as a standardized test to determine the specificity of the antibody binding, particularly in immunohistochemistry.

SPECIFIC IMMUNOLABELING TECHNIQUES

Immunolabeling for Light Microscopy

Light microscopy is the use of a light microscope, which is an instrument that requires the usage of light to view the enlarged specimen. In general, a compound light microscope is frequently used, where two lenses, the eyepiece, and the objective work simultaneously to generate the magnification of the specimen. Light microscopy frequently uses immunolabeling to observe targeted tissues or cells. For instance, a study was conducted to view the morphology and the production of hormones in pituitary adenoma cell cultures via light microscopy and other electron microscopic methods. This type of microscopy confirmed that the primary adenoma cell cultures keep their physiological characteristics *in vitro*, which matched the histology inspection. Moreover, cell cultures of human pituitary adenomas were viewed by light microscopy and immunocytochemistry, where these cells were fixed and immunolabeled with a monoclonal mouse antibody against human GH and a polyclonal

rabbit antibody against PRL. This is an example of how a immunolabeled cell culture of pituitary adenoma cells that were viewed via light microscopy and by other electron microscopy techniques can assist with the proper diagnosis of tumors.

Immunolabeling for Electron Microscopy

Electron microscopy (EM) is a focused area of science that uses the electron microscope as a tool for viewing tissues. Electron microscopy has a magnification level up to 2 million times, whereas light microscopy only has a magnification up to 1000-2000 times. There are two types of electron microscopes, the transmission electron microscope and the scanning electron microscope. Electron microscopy is a common method that uses the immunolabeling technique to view tagged tissues or cells. The electron microscope method follows many of the same concepts as immunolabeling for light microscopy, where the particular antibody is able to recognize the location of the antigen of interest and then be viewed by the electron microscope. The advantage of electron microscopy over light microscopy is the ability to view the targeted areas at their subcellular level. Generally, a heavy metal that is electron dense is used for EM, which can reflect the incident electrons. Immunolabeling is typically confirmed using the light microscope to assure the presence of the antigen and then followed up with the electron microscope. Immunolabeling and electron microscopy are often used to view chromosomes. A study was conducted to view possible improvements of immunolabeling chromosome structures, such as topoisomerase II α and condensin in dissected mitotic chromosomes. In particular, these investigators used UV irradiation of separated nuclei or showed how chromosomes assist by high levels of specific immunolabeling, which were viewed by electron microscopy.

Immunolabeling for Transmission Electron Microscopy

Transmission electron microscopy (TEM) uses a transmission electron microscope to form a two-dimensional image by shooting electrons through a thin piece of tissue. The brighter certain areas are on the image, the more electrons that are able to move through the specimen. Transmission Electron Microscopy has been used as a way to view immunolabeled tissues and cells. For instance, bacteria can be viewed by TEM when immunolabeling is applied. A study was conducted to examine the structures of CS3 and CS6 fimbriae in different *Escherichia coli* strains, which were detected by TEM followed by negative staining, and immunolabeling. More specifically, immunolabeling of the fimbriae confirmed the existence of different surface antigens.

Immunolabeling for Scanning Electron Microscopy

Scanning electron microscopy (SEM) uses a scanning electron microscope, which produces large images that are perceived as three-dimensional when, in fact, they are not. This type of microscope concentrates a beam of electrons across a very small area (2-3 nm) of the specimen in order to produce electrons from said specimen. These secondary electrons are detected by a sensor, and the image of the specimen is generated over a certain time period. Scanning Electron Microscopy is a frequently used immunolabeling technique. SEM is able to detect the surface of cellular components in high resolution. This immunolabeling technique is very similar to the immuno-fluorescence method, but a colloidal gold tag is used instead of a fluorophore. Overall, the concepts are very parallel in that an unconjugated primary antibody is used and sequentially followed by a tagged secondary antibody that works against the primary antibody. Sometimes SEM in conjunction with gold particle immunolabeling is troublesome in regards to the particles and charges resolution under the electron beam; however, this resolution setback has been resolved by the improvement of the SEM instrumentation by backscattered electron imaging. This is because electron backscattered diffraction patterns provide a clean surface of the sample to interact with the primary electron beam.

Immunolabeling with Gold (Immunogold Labeling)

Immunolabeling with gold particles, also known as immunogold staining, is used regularly with scanning electron microscopy and transmission electron microscopy to successfully identify the area within cells and tissues where antigens are located. The gold particle labeling technique was first published by Faulk, W. and Taylor, G. when they were able to tag gold particles to anti-salmonella rabbit gamma globulins in one step in order to identify the location of the antigens of salmonella. Studies have shown that the size of the gold particle must be enlarged (>40 nm) to view the cells in low magnification, but gold particles that are too large can decrease the efficiency of the binding of the gold tag. Scientists have concluded the usage of smaller gold particles (1-5 nm) should be enlarged and enhanced with silver. Although osmium tetroxide staining can scratch the silver, gold particle enhancement was found not to be susceptible to scratching by osmium tetroxide staining; therefore, many cell adhesion studies of different substrates can use the immunogold labeling mechanism via the enhancement of the gold particles.

INSTANTANEOUS WAVE-FREE RATIO

The instantaneous wave-free ratio (iFR, sometimes referred to as the instant wave-free ratio or instant flow reserve) is a diagnostic tool used to assess whether a stenosis is causing a limitation of blood flow in coronary arteries with subsequent ischemia. iFR is performed during cardiac catheterisation (angiography) using invasive coronary pressure wires which are placed in the coronary arteries that are to be assessed. Pressure wires are commonly used by interventional cardiologists to guide decisions to perform revascularization, either by stenting or bypass surgery.

ASSESSMENT OF CORONARY STENOSIS

Coronary artery blockages or stenoses that limit blood flow to the heart muscle can cause angina and can be treated by stenting or bypass surgery. Relief of a stenosis by stenting aims to restore vessel patency with improvement in blood flow leading to a reduction in angina symptoms. However, if stenoses are not flow limiting, then they can be safely left alone without stenting or surgery and this help reduce patient's exposure to unnecessary procedures and potential complications. Identifying stenoses that cause flow limitation, or ischaemia, can be done in a variety of ways. Non-invasive tests can include stress testing such as exercise electrocardiograms, stress echocardiography, or perfusion imaging tests such as scintigraphy or SPECT. Alternatively, invasive tests can be performed at the time of angiography, and these include those that measure coronary flow velocity in the vessel, CFR or index flow against pressure gradients such as hyperaemic or basal stenosis resistance (HSR or BSR). More commonly coronary pressure measurements are used as a surrogate for flow measurement and techniques include iFR and fractional flow reserve (FFR). Cardiologists use a combination of these investigations together with the patient's history, symptoms and clinical risk factors to decide if a stenosis requires further treatment.

IFR DEFINITION AND CALCULATION

Instantaneous wave-free ratio is performed using high fidelity pressure wires that are passed distal to the coronary stenosis. iFR isolates a specific period in diastole, called the wave-free period, and uses the ratio of distal coronary pressure (P_d) to the pressure observed in the aorta (P_a) over this period. During this wave-free period, the competing forces (waves) that affect coronary flow are quiescent meaning pressure and flow are linearly related as compared to the rest of the cardiac cycle. When stenoses are flow limiting, P_d and P_a pressures over the wave-free period diverge; a normal ratio is 1.0 and iFR values of below 0.90 suggest flow limitation. iFR can be calculated using dedicated consoles available for medical use and typically uses an average over 5

heart beats but can be performed using a single heart beat. iFR is measured at rest, without the need for pharmacological vasodilators or stressors and compares well to other invasive and non-invasive markers of ischemia or flow limitation.

STUDIES

A number of studies using the Imperial College developed iFR algorithm have been conducted.

The ADVISE study was a proof of concept study that demonstrated that the wave-free period, usually isolated using wave-intensity analysis, could be reliably determined using a pressure-only approach. This was shown across a variety of stenosis severities and demonstrated that over that specific period in diastole, microcirculatory resistance was the lowest and most stable compared to the rest of the cardiac cycle. During this specific period, pressure and flow are linearly related, allowing pressure-only inferences of transtenotic flow limitation. iFR was compared to the current clinically-used reference standard, FFR, and demonstrated good diagnostic lesion classification. Furthermore, iFR was consistent in utility for the left and right coronary artery, and the wave-free period was stable from beat-to-beat and even in the presence of arrhythmia. The study however was relatively underpowered to determine a clinical cut-point. The ADVISE-Registry was a larger study in which pressure wire data collected from patients undergoing functional assessment for clinical management, was analysed to compare iFR and FFR. This study recognised that matching FFR was limited by the capacity of FFR to match itself with repeated measures and accordingly presented per-range agreements. An independent consecutive blinded comparison of iFR and FFR in Asian patients reported similar results. The CLARIFY study was a physiological study in which iFR and FFR were compared to an independent third measure, hyperaemic stenosis resistance (HSR) which uses a combination of pressure and flow assessment to detect ischemia. Both measures compared equally well to this flow-based index. In stenoses classified as ischemic, the wave-free period provided a similar reduction in microvascular resistance as adenosine-mediated hyperaemia over the whole cardiac cycle. The Hybrid iFR-FFR approach has been proposed as method of minimising patient exposure to adenosine or other vasodilators. In this approach, an iFR can be measured, and stenoses with an iFR >0.93 may be deferred while those with an iFR <0.86 can be treated by revascularization. Stenoses with an iFR between 0.86-0.93 can undergo an FFR assessment to guide therapy. This range can be changed by the physician according to how closely they want to match an FFR-classification of lesions. Using this typical range, almost 60 per cent of stenoses can be spared from needing vasodilator infusion.

ADVISE-II has prospectively assessed the hybrid approach in an independent international multi-centre study which confirmed the findings of the earlier Hybrid analysis. Interim results were reported by Javier Escaned at EuroPCR, the final findings were presented at TCT 2013. These findings have been in keeping with an international collaborative study, RESOLVE, which pooled retrospective data from many centres worldwide with independent core-lab analysis of raw pressure traces. Conflicting data, when re-analysed using the clinically available iFR-algorithms produced results similar to the ADVISE-Registry and ADVISE Hybrid studies.

iFR has also been compared to FFR and the resting Pd/Pa ratio in the post-coronary intervention setting. In a manner similar to FFR, iFR can be used to measure the haemodynamic change induced by stent placement and detect the impact of any residual stenoses. The ADVISE-in-Practice study reported the first clinical experience of the commercially available iFR console. Video examples of iFR technology in clinical use are available here [1].

IFR CLINICAL OUTCOME STUDIES

Two outcome studies, DEFINE-FLAIR and the iFR-SWEDEHEART will assess whether patient outcomes differ if stenoses are treated according to iFR or FFR classification of stenosis severity. DEFINE-FLAIR, which aims to recruit 2500 patients, will be the largest physiological study to date and will provide the first

randomised data of the use of iFR and FFR in patients with both stable coronary disease and acute coronary syndromes. It will also provide the first randomised data of the clinical utility of FFR in guiding and deferring coronary intervention outside of the FAME and DEFER studies. Both DEFINE-FLAIR and iFR-SWEDEHEART have harmonised clinical endpoints which will enable combined analysis of over 4500 patients.

SYNTAX II is a modern PCI study assessing the clinical utility of stenting three-vessel coronary disease. Patients with three-vessel coronary disease whom the Heart Team decide are in equipoise between surgical revascularization or PCI intervention, undergo physiologically guided stenting. The SYNTAX II clinical risk calculator is used to determine equipoise. All major vessels undergo pressure wire assessment using the iFR-FFR hybrid approach, and only those positive for ischaemia undergo stenting. Intravascular ultrasound is used to ensure optimal stent delivery. Clinical outcomes will be compared to a historical surgical cohort from SYNTAX-I study. An example of a patient enrolled into the SYNTAX-II study is shown in this video.

IFR PHYSIOLOGICAL MAPPING OF CORONARY VESSELS

The instantaneous wave-free ratio can also be calculated on a beat-to-beat basis. If performed during the slow withdrawal of a pressure wire (referred to as ‘pullback’) under resting conditions, it is possible to plot the change of iFR or trans-stenotic gradient over the wave-free period throughout the vessel. This can help identify focal and diffuse coronary disease potentially amenable to revascularisation.. The pressure wire data can be co-registered with the angiographic findings to aid ease of interpretation. With mechanised pullback approaches which move the pressure wire at a fixed speed, the physiological length of a stenosis can be determined. With advanced live co-registration this is also feasible with during manual pullback. Advanced computer processing can utilise iFR-pullback data to predict the hemodynamic consequences of performing percutaneous coronary intervention (PCI) to a chosen segment of the coronary vessel. The predicted iFR values produced by the virtual-PCI on the iFR-pullback are closely related to those observed after PCI in the coronary vessel. This predictive capacity may provide added advantage to Interventional Cardiologists when assessing multiple coronary stenoses in a single vessel and planning the treatment approach. iFR-pullback technologies are undergoing evaluation in European centres.

CONTROVERSY

Alternative iFR computation systems have been proposed, for example incorporating part of systole into the definition of diastole and optionally different time-shifts between Pd and Pa signals, yielding unsatisfactory results. The same datasets reanalysed using the standard algorithms confirm the mainstream findings. Various explanations have been proposed.

LABORATORY DIAGNOSIS OF VIRAL INFECTIONS

In the diagnostic laboratory virus infections can be confirmed by a multitude of methods. Diagnostic virology has changed rapidly due to the advent of molecular techniques and increased clinical sensitivity of serological assays.

SAMPLING

A wide variety of samples can be used for virological testing. The type of sample sent to the laboratory often depends on the type of viral infection being diagnosed and the test required. Proper sampling technique is essential to avoid potential pre-analytical errors. For example, different types of samples must be collected in appropriate tubes to maintain the integrity of the sample and stored at appropriate temperatures (usually 4°C) to preserve the virus and prevent bacterial or fungal growth. Sometimes multiple sites may also be sampled.

Types of samples include:

- Blood
- Skin
- Sputum, gargles and bronchial washings
- Urine
- Semen
- Faeces
- Cerebrospinal fluid
- Tissues (biopsies or post-mortem)
- Dried blood spots

VIRUS ISOLATION

Viruses are often isolated from the initial patient sample. This allows the virus sample to be grown into larger quantities and allows a larger number of tests to be run on them. This is particularly important for samples that contain new or rare viruses for which diagnostic tests are not yet developed. Many viruses can be grown in cell culture in the lab. To do this, the virus sample is mixed with cells, a process called adsorption, after which the cells become infected and produce more copies of the virus. Although different viruses often only grow in certain types of cells, there are cells that support growth of a large variety of viruses and are a good starting point, for example the African monkey kidney cell line (Vero cells), human lung fibroblasts (MRC-5), and human epidermoid carcinoma cells (HEp-2). One sign of knowing whether the cells are successfully replicating the virus is to check for a change in cell morphology or for the presence of cell death using a microscope. Other viruses may require alternative methods for growth such as the inoculation of embryonated chicken eggs (*e.g.* avian influenza viruses) or the intracranial inoculation of virus using newborn mice (*e.g.* lyssaviruses).

NUCLEIC ACID BASED METHODS

Molecular techniques are the most specific and sensitive diagnostic tests. They are capable of detecting either the whole viral genome or parts of the viral genome. In the past nucleic acid tests have mainly been used as a secondary test to confirm positive serological results. However, as they become cheaper and more automated, they are increasingly becoming the primary tool for diagnostics.

Polymerase Chain Reaction

Detection of viral RNA and DNA genomes can be performed using polymerase chain reaction. This technique makes many copies of the virus genome using virus-specific probes. Variations of PCR such as nested reverse transcriptase PCR and real time PCR can also be used to determine viral loads in patient serum. This is often used to monitor treatment success in HIV cases.

Sequencing

Sequencing is the only diagnostic method that will provide the full sequence of a virus genome. Hence, it provides the most information about very small differences between two viruses that would look the same using other diagnostic tests. Currently it is only used when this depth of information is required. For example, sequencing is useful when specific mutations in the patient are tested for in order to determine antiviral therapy and susceptibility to infection. However, as the tests are getting cheaper, faster and more automated, sequencing will likely become the primary diagnostic tool in the future.

MICROSCOPY BASED METHODS

Immunofluorescence or Immunoperoxidase

Immunofluorescence or immunoperoxidase assays are commonly used to detect whether a virus is present in a tissue sample. These tests are based on the principle that if the tissue is infected with a virus, an antibody specific to that virus will be able to bind to it. To do this, antibodies that are specific to different types of viruses are mixed with the tissue sample. After the tissue is exposed to a specific wavelength of light or a chemical that allows the antibody to be visualized. These tests require specialized antibodies that are produced and purchased from commercial companies. These commercial antibodies are usually well characterized and are known to bind to only one specific type of virus. They are also conjugated to a special kind of tag that allows the antibody to be visualized in the lab, i.e. so that it will emit fluorescence or a colour. Hence, immunofluorescence refers to the detection of a fluorescent antibody (immuno) and immunoperoxidase refers to the detection of a colored antibody (peroxidase produces a dark brown colour).

HOST ANTIBODY DETECTION

A person who has recently been infected by a virus will produce antibodies in their bloodstream that specifically recognize that virus. This is called humoral immunity. Two types of antibodies are important. The first called IgM is highly effective at neutralizing viruses but is only produced by the cells of the immune system for a few weeks.

The second, called, IgG is produced indefinitely. Therefore, the presence of IgM in the blood of the host is used to test for acute infection, whereas IgG indicates an infection sometime in the past. Both types of antibodies are measured when tests for immunity are carried out. Antibody testing has become widely available. It can be done for individual viruses (*e.g.* using an ELISA assay) but in automated panels that can screen for many viruses at once are becoming increasingly common.

HEMAGGLUTINATION ASSAY

Some viruses attach to molecules present on the surface of red blood cells, for example, influenza virus. A consequence of this is that – at certain concentrations – a viral suspension may bind together (agglutinate) the red blood cells thus preventing them from settling out of suspension.

MARTHA MITCHELL EFFECT

The Martha Mitchell effect is the process by which a psychiatrist, psychologist, or other mental health clinician labels the patient's accurate perception of real events as delusional and misdiagnoses accordingly.

DESCRIPTION

According to Bell et al., “Sometimes, improbable reports are erroneously assumed to be symptoms of mental illness”, due to a “failure or inability to verify whether the events have actually taken place, no matter how improbable intuitively they might appear to the busy clinician”. They note that typical examples of such situations, may include:

- Pursuit by organized criminals
- Surveillance by law enforcement officers
- Infidelity by a spouse
- Physical issues

Quoting psychotherapist Joseph Berke, the authors note that “even paranoids have enemies”. Any patient, they explain, can be misdiagnosed by clinicians, especially patients with a history of paranoid delusions. Of note is how habitually patients are diagnosed as delusional when their grievances concern health care workers and/or health care institutions, even when the patient has no history of delusion. “A patient arriving claiming to have been injured by another health care professional is regarded as a crazy person who potentially could ruin the career of an innocent colleague.”

ORIGIN

Psychologist Brendan Maher named the effect after Martha Beall Mitchell. Mrs. Mitchell was the wife of John Mitchell, Attorney-General in the Nixon administration. When she alleged that White House officials were engaged in illegal activities, her claims were attributed to mental illness. Ultimately, however, the facts of the Watergate scandal vindicated her and garnered her the label, “The Cassandra of Watergate”. However, many of her allegations remain unproven, such as her claim that she had been drugged and put under guard during a visit to California after her husband had been summoned back to Washington, D.C. in order to prevent her from leaving the hotel or making phone calls to the news media.

MEDICAL ERROR

A medical error is a preventable adverse effect of care, whether or not it is evident or harmful to the patient. This might include an inaccurate or incomplete diagnosis or treatment of a disease, injury, syndrome, behaviour, infection, or other ailment. Globally, it is estimated that 142,000 people died in 2013 from adverse effects of medical treatment; this is an increase from 94,000 in 1990. However, a 2016 study of the number of deaths that were a result of medical error in the U.S. placed the yearly death rate in the U.S. alone at 251,454 deaths, which suggests that the 2013 global estimation may not be accurate.

DEFINITIONS

The word *error* in medicine is used as a label for nearly all of the clinical incidents that harm patients. Medical errors are often described as human errors in health care. Whether the label is a medical error or human error, one definition used in medicine says that it occurs when a health care provider chooses an inappropriate method of care, improperly executes an appropriate method of care, or reads the wrong CT scan. It has been said that the definition should be the subject of more debate. For instance, studies of hand hygiene compliance of physicians in an ICU show that compliance varied from 19 per cent to 85 per cent. The deaths that result from infections caught as a result of treatment providers improperly executing an appropriate method of care by not complying with known safety standards for hand hygiene are difficult to regard as innocent accidents or mistakes. At the least, they are negligence, if not dereliction, but in medicine they are lumped together under the word *error* with innocent accidents and treated as such. There are many types of medical error, from minor to major, and causality is often poorly determined. There are many taxonomies for classifying medical errors.

IMPACT

Globally, it is estimated that 142,000 people died in 2013 from adverse effects of medical treatment; in 1990, the number was 94,000. A 2000 Institute of Medicine report estimated that medical errors result in between 44,000 and 98,000 preventable deaths and 1,000,000 excess injuries each year in U.S. hospitals. In the UK, a 2000 study found that an estimated 850,000 medical errors occur each year, costing over £2 billion. Some researchers questioned the accuracy of the IOM study, criticizing the statistical handling of measurement errors in the report, significant subjectivity in determining which deaths were “avoidable” or due to medical error, and

an erroneous assumption that 100 per cent of patients would have survived if optimal care had been provided. A 2001 study in the *Journal of the American Medical Association* of seven Department of Veterans Affairs medical centers estimated that for roughly every 10,000 patients admitted to the select hospitals, one patient died who would have lived for three months or more in good cognitive health had “optimal” care been provided. A 2006 follow-up to the IOM study found that medication errors are among the most common medical mistakes, harming at least 1.5 million people every year. According to the study, 400,000 preventable drug-related injuries occur each year in hospitals, 800,000 in long-term care settings, and roughly 530,000 among Medicare recipients in outpatient clinics. The report stated that these are likely to be conservative estimates. In 2000 alone, the extra medical costs incurred by preventable drug-related injuries approximated \$887 million—and the study looked only at injuries sustained by Medicare recipients, a subset of clinic visitors. None of these figures take into account lost wages and productivity or other costs. According to a 2002 Agency for Health care Research and Quality report, about 7,000 people were estimated to die each year from medication errors - about 16 percent more deaths than the number attributable to work-related injuries (6,000 deaths). Medical errors affect one in 10 patients worldwide. One extrapolation suggests that 180,000 people die each year partly as a result of iatrogenic injury. One in five Americans (22 per cent) report that they or a family member have experienced a medical error of some kind. The World Health Organization registered 14 million new cases and 8.2 million cancer-related deaths in 2012. It estimated that the number of cases could increase by 70 per cent through 2032. As the number of cancer patients receiving treatment increases, hospitals around the world are seeking ways to improve patient safety, to emphasize traceability and raise efficiency in their cancer treatment processes.

A study released in 2016 found medical error is the third leading cause of death in the United States, after heart disease and cancer. Researchers looked at studies that analyzed the medical death rate data from 2000 to 2008 and extrapolated that over 250,000 deaths per year had stemmed from a medical error, which translates to 9.5 per cent of all deaths annually in the US.

Difficulties in Measuring Frequency of Errors

About 1 per cent of hospital admissions result in an adverse event due to negligence. However, mistakes are likely much more common, as these studies identify only mistakes that led to measurable adverse events occurring soon after the errors. Independent review of doctors’ treatment plans suggests that decision-making could be improved in 14 per cent of admissions; many of the benefits would have delayed manifestations. Even this number may be an underestimate. One study suggests that adults in the United States receive only 55 per cent of recommended care. At the same time, a second study found that 30 per cent of care in the United States may be unnecessary. For example, if a doctor fails to order a mammogram that is past due, this mistake will not show up in the first type of study. In addition, because no adverse event occurred during the short follow-up of the study, the mistake also would not show up in the second type of study because only the principal treatment plans were critiqued. However, the mistake would be recorded in the third type of study. If a doctor recommends an unnecessary treatment or test, it may not show in any of these types of studies. Cause of death on United States death certificates, statistically compiled by the Centers for Disease Control and Prevention (CDC), are coded in the International Classification of Disease (ICD), which does not include codes for human and system factors.

CAUSES

Medical errors are associated with inexperienced physicians and nurses, new procedures, extremes of age, and complex or urgent care. Poor communication (whether in one’s own language or, as may be the case for medical tourists, another language), improper documentation, illegible handwriting, inadequate nurse-to-patient ratios, and similarly named medications are also known to contribute to the problem. Patient actions may also contribute significantly to medical errors. Falls, for example, may result from patients’ own misjudgements.

Human error has been implicated in nearly 80 percent of adverse events that occur in complex health care systems. The vast majority of medical errors result from faulty systems and poorly designed processes versus poor practices or incompetent practitioners.

Health care Complexity

Complicated technologies, powerful drugs, intensive care, and prolonged hospital stay can contribute to medical errors.

System and Process Design

In 2000, The Institute of Medicine released “To Err is Human,” which asserted that the problem in medical errors is not bad people in health care—it is that good people are working in bad systems that need to be made safer. Poor communication and unclear lines of authority of physicians, nurses, and other care providers are also contributing factors. Disconnected reporting systems within a hospital can result in fragmented systems in which numerous hand-offs of patients results in lack of coordination and errors. Other factors include the impression that action is being taken by other groups within the institution, reliance on automated systems to prevent error, and inadequate systems to share information about errors, which hampers analysis of contributory causes and improvement strategies. Cost-cutting measures by hospitals in response to reimbursement cutbacks can compromise patient safety. In emergencies, patient care may be rendered in areas poorly suited for safe monitoring. The American Institute of Architects has identified concerns for the safe design and construction of health care facilities. Infrastructure failure is also a concern. According to the WHO, 50 per cent of medical equipment in developing countries is only partly usable due to lack of skilled operators or parts. As a result, diagnostic procedures or treatments cannot be performed, leading to substandard treatment.

The Joint Commission’s Annual Report on Quality and Safety 2007 found that inadequate communication between health care providers, or between providers and the patient and family members, was the root cause of over half the serious adverse events in accredited hospitals. Other leading causes included inadequate assessment of the patient’s condition, and poor leadership or training.

Competency, Education, and Training

Variations in health care provider training and experience and failure to acknowledge the prevalence and seriousness of medical errors also increase the risk. The so-called July effect occurs when new residents arrive at teaching hospitals, causing an increase in medication errors according to a study of data from 1979-2006.

Human Factors and Ergonomics

Cognitive errors commonly encountered in medicine were initially identified by psychologists Amos Tversky and Daniel Kahneman in the early 1970s. Jerome Groopman, author of *How Doctors Think*, says these are “cognitive pitfalls”, biases which cloud our logic. For example, a practitioner may overvalue the first data encountered, skewing his thinking (or recent or dramatic cases which come quickly to mind and may colour judgement). Another pitfall is where stereotypes may prejudice thinking. Sleep deprivation has also been cited as a contributing factor in medical errors. One study found that being awake for over 24 hours caused medical interns to double or triple the number of preventable medical errors, including those that resulted in injury or death. The risk of car crash after these shifts increased by 168 per cent, and the risk of near miss by 460 per cent. Interns admitted falling asleep during lectures, during rounds, and even during surgeries. Night shifts are associated with worse surgeon performance during laparoscopic surgeries.

Practitioner risk factors include fatigue, depression, and burnout. Factors related to the clinical setting include diverse patients, unfamiliar settings, time pressures, and increased patient-to-nurse staffing ratio increases. Drug names that look alike or sound alike are also a problem.

EXAMPLES

Errors can include misdiagnosis or delayed diagnosis, administration of the wrong drug to the wrong patient or in the wrong way, giving multiple drugs that interact negatively, surgery on an incorrect site, failure to remove all surgical instruments, failure to take the correct blood type into account, or incorrect record-keeping. A 10th type of error is ones which are not watched for by researchers, such as RNs failing to programme an IV pump to give a full dose of IV antibiotics or other medication.

Errors in Diagnosis

A large study reported several cases where patients were wrongly told that they were HIV-negative when the physicians erroneously ordered and interpreted HTLV (a closely related virus) testing rather than HIV testing. In the same study, >90 per cent of HTLV tests were ordered erroneously. It is estimated that between 10-15 percent of physician diagnoses are erroneous. Misdiagnosis of lower extremity cellulitis is estimated to occur in 30 per cent of patients, leading to unnecessary hospitalizations in 85 per cent and unnecessary antibiotic use in 92 per cent. Collectively, these errors lead to between 50,000 and 130,000 unnecessary hospitalizations and between \$195 and \$515 million in avoidable health care spending annually in the United States.

Misdiagnosis of Psychological Disorders

Female sexual desire sometimes used to be diagnosed as female hysteria. Sensitivities to foods and food allergies risk being misdiagnosed as the anxiety disorder Orthorexia. Studies have found that bipolar disorder has often been misdiagnosed as major depression. Its early diagnosis necessitates that clinicians pay attention to the features of the patient's depression and also look for present or prior hypomanic or manic symptomatology. The misdiagnosis of schizophrenia is also a common problem. There may be long delays of patients getting a correct diagnosis of this disorder. The DSM-5 field trials included "test-retest reliability" which involved different clinicians doing independent evaluations of the same patient—a new approach to the study of diagnostic reliability.

Outpatient vs. Inpatient

Misdiagnosis is the leading cause of medical error in outpatient facilities. Since the National Institute of Medicine's 1999 report, "To Err is Human," found up to 98,000 hospital patients die from preventable medical errors in the U.S. each year, government and private sector efforts have focused on inpatient safety.

AFTER AN ERROR HAS OCCURRED

Mistakes can have a strongly negative emotional impact on the doctors who commit them.

Recognizing that Mistakes are not Isolated Events

Some physicians recognize that adverse outcomes from errors usually do not happen because of an isolated error and actually reflect system problems. This concept is often referred to as the Swiss Cheese Model. This is the concept that there are layers of protection for clinicians and patients to prevent mistakes from occurring. Therefore, even if a doctor or nurse makes a small error (e.g. incorrect dose of drug written on a drug chart by doctor), this is picked up before it actually affects patient care (e.g. pharmacist checks the drug chart and

rectifies the error). Such mechanisms include: Practical alterations (*e.g.* -medications that cannot be given through IV, are fitted with tubing which means they cannot be linked to an IV even if a clinician makes a mistake and tries to), systematic safety processes (*e.g.* all patients must have a Waterlow score assessment and falls assessment completed on admission), and training programmes/continuing professional development courses are measures that may be put in place.

There may be several breakdowns in processes to allow one adverse outcome. In addition, errors are more common when other demands compete for a physician's attention. However, placing too much blame on the system may not be constructive.

Placing the Practice of Medicine in Perspective

Essayists imply that the potential to make mistakes is part of what makes being a physician rewarding and without this potential the rewards of medical practice would be diminished. Laurence states that "Everybody dies, you and all of your patients. All relationships end. Would you want it any other way? [...] Don't take it personally" Seder states "[...] if I left medicine, I would mourn its loss as I've mourned the passage of my poetry. On a daily basis, it is both a privilege and a joy to have the trust of patients and their families and the camaraderie of peers. There is no challenge to make your blood race like that of a difficult case, no mind game as rigorous as the challenging differential diagnosis, and though the stakes are high, so are the rewards."

Disclosing Mistakes

Forgiveness, which is part of many cultural traditions, may be important in coping with medical mistakes.

To Oneself

Inability to forgive oneself may create a cycle of distress and increased likelihood of a future error.

However, Wu et al. suggest "...those who coped by accepting responsibility were more likely to make constructive changes in practice, but [also] to experience more emotional distress." It may be helpful to consider the much larger number of patients who are not exposed to mistakes and are helped by medical care.

To Patients

Gallagher et al. state that patients want "information about what happened, why the error happened, how the error's consequences will be mitigated, and how recurrences will be prevented." Interviews with patients and families reported in a 2003 book by Rosemary Gibson and Janardan Prasad Singh, put forward that those who have been harmed by medical errors face a "wall of silence" and "want an acknowledgement" of the harm. With honesty, "healing can begin not just for the patients and their families but also the doctors, nurses and others involved." Detailed suggestions on how to disclose are available. A 2005 study by Wendy Levinson of the University of Toronto showed surgeons discussing medical errors used the word "error" or "mistake" in only 57 percent of disclosure conversations and offered a verbal apology only 47 percent of the time. Patient disclosure is important in the medical error process. The current standard of practice at many hospitals is to disclose errors to patients when they occur. In the past, it was a common fear that disclosure to the patient would incite a malpractice lawsuit. Many physicians would not explain that an error had taken place, causing a lack of trust towards the health care community. In 2007, 34 states passed legislation that precludes any information from a physician's apology for a medical error from being used in malpractice court (even a full admission of fault). This encourages physicians to acknowledge and explain mistakes to patients, keeping an open line of communication.

The American Medical Association's Council on Ethical and Judicial Affairs states in its ethics code:

- “Situations occasionally occur in which a patient suffers significant medical complications that may have resulted from the physician's mistake or judgement. In these situations, the physician is ethically required to inform the patient of all facts necessary to ensure understanding of what has occurred. Concern regarding legal liability which might result following truthful disclosure should not affect the physician's honesty with a patient.”

From the American College of Physicians Ethics Manual:

- “In addition, physicians should disclose to patients information about procedural or judgement errors made in the course of care if such information is material to the patient's well-being. Errors do not necessarily constitute improper, negligent, or unethical behaviour, but failure to disclose them may.”

However, “there appears to be a gap between physicians' attitudes and practices regarding error disclosure. Willingness to disclose errors was associated with higher training level and a variety of patient-centered attitudes, and it was not lessened by previous exposure to malpractice litigation”. Hospital administrators may share these concerns. Consequently, in the United States, many states have enacted laws excluding expressions of sympathy after accidents as proof of liability. However, “excluding from admissibility in court proceedings apologetic expressions of sympathy but not fault-admitting apologies after accidents”

Disclosure may actually reduce malpractice payments.

To Non-physicians

In a study of physicians who reported having made a mistake, it was offered that disclosing to non-physician sources of support may reduce stress more than disclosing to physician colleagues. This may be due to the finding that of the physicians in the same study, when presented with a hypothetical scenario of a mistake made by another colleague, only 32 per cent of them would have unconditionally offered support. It is possible that greater benefit occurs when spouses are physicians.

To other Physicians

Discussing mistakes with other physicians is beneficial. However, medical providers may be less forgiving of one another. The reason is not clear, but one essayist has admonished, “Don't Take Too Much Joy in the Mistakes of Other Doctors.”

To the Physician's Institution

Disclosure of errors, especially ‘near misses’ may be able to reduce subsequent errors in institutions that are capable of reviewing near misses. However, doctors report that institutions may not be supportive of the doctor.

Use of Rationalization to Cover up Medical Errors

Based on anecdotal and survey evidence, Banja states that rationalization (making excuses) is very common among the medical profession to cover up medical errors.

By Presence of the Patient

A survey of more than 10,000 physicians in the United States came to the results that, on the question “Are there times when it's acceptable to cover up or avoid revealing a mistake if that mistake would not cause harm to the patient?”, 19 per cent answered yes, 60 per cent answered no and 21 per cent answered it depends. On the

question “Are there times when it is acceptable to cover up or avoid revealing a mistake if that mistake would potentially or likely harm the patient?”, 2 per cent answered *yes*, 95 per cent answered *no* and 3 per cent answered *it depends*.

Cause-specific Preventive Measures

Traditionally, errors are attributed to mistakes made by individuals who may be penalized for these mistakes. The usual approach to correct the errors is to create new rules with additional checking steps in the system, aiming to prevent further errors. As an example, an error of free flow IV administration of heparin is approached by teaching staff how to use the IV systems and to use special care in setting the IV pump. While overall errors become less likely, the checks add to workload and may in themselves be a cause of additional errors. A newer model for improvement in medical care takes its origin from the work of W. Edwards Deming in a model of Total Quality Management. In this model, there is an attempt to identify the underlying system defect that allowed the opportunity for the error to occur. As an example, in such a system the error of free flow IV administration of Heparin is dealt with by not using IV heparin and substituting subcutaneous administration of heparin, obviating the entire problem. However, such an approach presupposes available research showing that subcutaneous heparin is as effective as IV. Thus, most systems use a combination of approaches to the problem.

In Specific Specialties

The field of medicine that has taken the lead in systems approaches to safety is anaesthesiology. Steps such as standardization of IV medications to 1 ml doses, national and international colour-coding standards, and development of improved airway support devices has made anesthesia care a model of systems improvement in care. Pharmacy professionals have extensively studied the causes of errors in the prescribing, preparation, dispensing and administration of medications. As far back as the 1930s, pharmacists worked with physicians to select, from many options, the safest and most effective drugs available for use in hospitals. The process is known as the Formulary System and the list of drugs is known as the Formulary. In the 1960s, hospitals implemented unit dose packaging and unit dose drug distribution systems to reduce the risk of wrong drug and wrong dose errors in hospitalized patients; centralized sterile admixture services were shown to decrease the risks of contaminated and infected intravenous medications; and pharmacists provided drug information and clinical decision support directly to physicians to improve the safe and effective use of medications. Pharmacists are recognized experts in medication safety and have made many contributions that reduce error and improve patient care over the last 50 years. More recently, governments have attempted to address issues like patient-pharmacists communication and consumer knowledge through measures like the Australian Government’s Quality Use of Medicines policy.

Legal Procedure

Standards and regulations for medical malpractice vary by country and jurisdiction within countries. Medical professionals may obtain professional liability insurances to offset the risk and costs of lawsuits based on medical malpractice.

PREVENTION

Medical care is frequently compared adversely to aviation; while many of the factors that lead to errors in both fields are similar, aviation’s error management protocols are regarded as much more effective. Safety measures include informed consent, the availability of a second practitioner’s opinion, voluntary reporting of errors, root cause analysis, reminders to improve patient medication adherence, hospital accreditation, and

systems to ensure review by experienced or specialist practitioners. A template has been developed for the design (both structure and operation) of hospital medication safety programmes, particularly for acute tertiary settings, which emphasizes safety culture, infrastructure, data (error detection and analysis), communication and training. Particularly to prevent the medication errors in the perspective of the intrathecal administration of local anaesthetics, there is a proposal to change the presentation and packaging of the appliances and agents used for this purpose. One spinal needle with a syringe prefilled with the local anaesthetic agents may be marketed in a single blister pack, which will be peeled open and presented before the anaesthesiologist conducting the procedure.

Reporting Requirements

In the United States, adverse medical event reporting systems were mandated in just over half (27) of the states as of 2014, a figure unchanged since 2007. In U.S. hospitals error reporting is a condition of payment by Medicare. An investigation by the Office of Inspector General, Department of Health and Human Services released January 6, 2012 found that most errors are not reported and even in the case of errors that are reported and investigated changes are seldom made which would prevent them in the future. The investigation revealed that there was often lack of knowledge regarding which events were reportable and recommended that lists of reportable events be developed.

MISCONCEPTIONS

These are the common misconceptions about adverse events, and the arguments and explanations against those misconceptions are noted in parentheses:

- “Bad apples” or incompetent health care providers are a common cause. (Although human error is commonly an initiating event, the faulty process of delivering care invariably permits or compounds the harm, and is the focus of improvement.)
- High risk procedures or medical specialties are responsible for most *avoidable* adverse events. (Although some mistakes, such as in surgery, are harder to conceal, errors occur in all levels of care. Even though complex procedures entail more risk, adverse outcomes are not usually due to error, but to the severity of the condition being treated.). However, USP has reported that medication errors during the course of a surgical procedure are three times more likely to cause harm to a patient than those occurring in other types of hospital care.)
- If a patient experiences an adverse event during the process of care, an error has occurred. (Most medical care entails some level of risk, and there can be complications or side effects, even unforeseen ones, from the underlying condition or from the treatment itself.)

MEDICAL FINDINGS

Medical findings will signify the collective physical and psychological occurrences of patients surveyed by a medical doctor. The survey is composed of physical examinations by the doctor’s senses and simple medical devices, which build clinical findings. If necessary, the results are proofed by further diagnostic tests, which may include procedures using medical apparatus. As the findings relate to the intersubjective occurrences diagnosed by the doctor, they will be differentiated from the sum of subjective data in the patient’s medical history.

DEFINITION

Using for example the National Cancer Institute Thesaurus (NCIT), a finding is a Clinical, Laboratory or Molecular evidence (or absence of evidence) of a disease.

TYPES OF FINDINGS

The documentation of medical findings can differ between:

- Quantitative evidence, e.g. “the cholesterol level is 220 mmol/L”
- Qualitative evidence, e.g. “the consciousness is clouded”

Out of a combination of these extracted medical findings, a doctor can deduce their diagnosis, whereby they ascribe an affliction to the patient. Medical findings that do not match these diagnoses will be labeled auxiliary medical findings, whose symptoms can illustrate other afflictions.

According to the nature of the finding they can be classified into:

- *Clinical:* When they are related to the patient’s medical signs and symptoms evolution
 - *Physiological:* When they are related to an intermediate biological biomarker
 - *Pathological or histopathological:* When they are related to the physical damage produced by the disease
- Psychological findings would be more a medical sign than a finding except when they are supported by a test. Findings can appear during a targeted medical test or unexpectedly during an unrelated exploration.

FINDINGS AND SYMPTOMS

Although the difference is not clear in everyday parlance, medical findings and symptoms are completely delineated: during the process of confirming a medical finding the empirical character (those that can be found) of the collected medical characteristics is brought out, highlighting symptom of an affliction. Acquiring medical findings is one of the main duties of a doctor.

MEDICAL LABORATORY

A medical laboratory or clinical laboratory is a laboratory where tests are usually done on clinical specimens in order to obtain information about the health of a patient as pertaining to the diagnosis, treatment, and prevention of disease. Clinical laboratories are thus focused on applied science mainly on a production-like basis, as opposed to research laboratories that focus on basic science on an academic basis.

DEPARTMENTS

Laboratory medicine is generally divided into two sections, each of which being subdivided into multiple units. These two sections are:

- Anatomic pathology: units included here are histopathology, cytopathology, and electron microscopy. Academically, each unit is studied alone in one course. Other courses pertaining to this section include anatomy, physiology, histology, pathology, and pathophysiology.
- Clinical pathology, which includes:
 - (a) *Clinical Microbiology:* This encompasses five different sciences (units). These include bacteriology, virology, parasitology, immunology, and mycology.
 - (b) *Clinical Chemistry:* Units under this busy section include instrumental analysis of blood components, enzymology, toxicology and endocrinology.
 - (c) *Hematology:* This section consists of automated and manual analysis of blood cells. It includes two subunits, which are coagulation and blood bank.
 - (d) Genetics is also studied along with a subspecialty known as cytogenetics.
 - (e) *Reproductive biology:* Semen analysis, Sperm bank and assisted reproductive technology.

Distribution of clinical laboratories in health institutions varies greatly from one place to another. For instance, for microbiology, some health facilities have a single laboratory for microbiology, while others have a separate lab for each unit, with nothing called a “microbiology” lab.

The following is a detailed breakdown of the responsibilities of each unit:

- Microbiology receives almost any clinical specimen, including swabs, feces, urine, blood, sputum, cerebrospinal fluid, synovial fluid, as well as possible infected tissue. The work here is mainly concerned with cultures, to look for suspected pathogens which, if found, are further identified based on biochemical tests. Also, sensitivity testing is carried out to determine whether the pathogen is sensitive or resistant to a suggested medicine. Results are reported with the identified organism(s) and the type and amount of drug(s) that should be prescribed for the patient.
- Parasitology responsible for examining parasites, feces samples may be examined for evidence of intestinal parasites such as tapeworms or hookworms and others.
- Virology is concerned with identification of viruses in specimens such as blood, urine, and cerebrospinal fluid.
- Hematology works with whole blood to do full blood counts, and blood films as well as many other specialised tests.
- Coagulation requires citrated blood samples to analyze blood clotting times and coagulation factors.
- Clinical Biochemistry usually receives serum or plasma. They test the serum for chemicals present in blood. These include a wide array of substances, such as lipids, blood sugar, enzymes, and hormones.
- Toxicology mainly tests for pharmaceutical and recreational drugs. Urine and blood samples are submitted to this lab.
- Immunology/Serology uses the concept of antigen-antibody interaction as a diagnostic tool. Compatibility of transplanted organs is also determined.
- Immunohaematology, or Blood bank determines blood groups, and performs compatibility testing on donor blood and recipients. It also prepares blood components, derivatives, and products for transfusion. Regulated by the FDA since giving blood is considered a drug, this unit determines a patient's blood type and Rh status, checks for antibodies to common antigens found on red blood cells, and cross matches units that are negative for the antigen.
- Urinalysis tests urine for many analytes. Some health care providers have a urinalysis laboratory, while others don't. Instead, each component of the urinalysis is performed at the corresponding unit. If measuring urine chemicals is required, the specimen is processed in the clinical biochemistry lab, but if cell studies are indicated, the specimen should be submitted to the cytopathology lab, and so on.
- Histopathology processes solid tissue removed from the body (biopsies) for evaluation at the microscopic level.
- Cytopathology examines smears of cells from all over the body (such as from the cervix) for evidence of inflammation, cancer, and other conditions.
- Electron microscopy prepares specimens and takes micrographs of very fine details by means of TEM and SEM.
- Genetics mainly performs DNA analysis.
- Cytogenetics involves using blood and other cells to get a karyotype. This can be helpful in prenatal diagnosis (*e.g.* Down's syndrome) as well as in cancer (some cancers have abnormal chromosomes).
- Surgical pathology examines organs, limbs, tumors, fetuses, and other tissues biopsied in surgery such as breast mastectomies.

MEDICAL LABORATORY STAFF

The staff of clinical laboratories may include:

- Pathologist
- Clinical Biochemist

- Pathologists' Assistant (PA)
- Biomedical Scientist (BMS) in the UK, Medical Laboratory Scientist (MT, MLS or CLS) in the US or Medical Laboratory Technologist in Canada
- Medical Laboratory Technician (MLT in US)
- Medical Laboratory Assistant (MLA)
- Phlebotomist (PBT)
- Histotechnologist/Histology Technician

TYPES OF LABORATORY

In many countries, there are two main types of labs that process the majority of medical specimens. Hospital laboratories are attached to a hospital, and perform tests on patients. Private (or community) laboratories receive samples from general practitioners, insurance companies, clinical research sites and other health clinics for analysis. For extremely specialised tests, samples may go to a research laboratory. A lot of samples are sent between different labs for uncommon tests. It is more cost effective if a particular laboratory specializes in a rare test, receiving specimens (and money) from other labs, while sending away tests it cannot do.

In many countries there are mainly three types of Medical Laboratories as per the types of investigations carried out.

- Clinical Pathology: Haematology, Histopathology, Cytology, Routine Pathology
- Clinical Microbiology: Bacteriology, Mycobacteriology, Virology, Mycology, Parasitology, Immunology, Serology.
- Clinical Biochemistry: Biochemical analysis, Hormonal assays etc. Blood Banks:- Blood bank is a separate body. Its laboratory need Microbiological analysis for infectious diseases that may be found in blood. Pathology to observe Blood grouping, Haematology and cross matching reactions. It also involves PRO department for the communication and contact for blood donations etc.. Molecular diagnostic lab or cytogenetics and molecular biology lab is the latest addition to the three types of medical laboratories listed above in many countries.

SPECIMEN PROCESSING AND WORK FLOW

Sample processing will usually start with a set of samples and a request form.

Typically a set of vacutainer tubes containing blood, or any other specimen, will arrive to the laboratory in a small plastic bag, along with the form. The form and the specimens are given a laboratory number. The specimens will usually all receive the same number, often as a sticker that can be placed on the tubes and form. This label has a barcode that can be scanned by automated analyzers and test requests uploaded from the LIS. Entry of requests onto a laboratory management system involves typing, or scanning (where barcodes are used) in the laboratory number, and entering the patient identification, as well as any tests requested. This allows laboratory machines, computers and staff to know what tests are pending, and also gives a place (such as a hospital department, doctor or other customer) for results to go. For biochemistry samples, blood is usually centrifuged and serum is separated. If the serum needs to go on more than one machine, it can be divided into separate tubes. Many specimens end up in one or more sophisticated automated analysers, that process a fraction of the sample and return one or more "results". Some laboratories use robotic sample handlers (Laboratory automation) to optimize the workflow and reduce contamination risk and sample handling of the staff. The work flow in a lab is usually heavy from 2:00 am to 10:00 am. Nurses and doctors generally have their patients tested at least once a day with general complete blood counts and chemistry profiles. These orders are then drawn during a morning run by phlebotomists for results to be available in the patient's charts for the attending physicians to consult during their morning rounds. Another busy time for the lab is after 3:00 pm when private

practice physician offices are closing. Couriers will pick up specimens that have been drawn throughout the day and deliver them to the lab. Also, couriers will stop at outpatient drawing centers and pick up specimens. These specimens will be processed in the evening and overnight to ensure results will be available the following day.

LABORATORY INFORMATICS

Laboratories today are held together by a system of software programmes, computers, and terminology standards that exchange data about patients, test requests, and test results known as a Laboratory information system or LIS. The LIS is often interfaced with the hospital information system, EHR and/or Laboratory instruments. Reporting of tests and results are being standardized to terminologies like Logical Observation Identifiers Names and Codes (LOINC) and Nomenclature for Properties and Units terminology (NPU terminology). This system enables hospitals and labs to order the correct test requests for each patient, keep track of individual patient or specimen histories, and help guarantee a better quality of results as well as printing hard copies of the results for patient charts and doctors to check.

RESULT ANALYSIS, VALIDATION AND INTERPRETATION

According to ISO 15189 norm, all pathological results must be verified by a competent professional. In some countries staff like clinical scientists do the majority of this work inside the laboratory with abnormal results referred to the relevant pathologist. In others, only medical staff (pathologist or clinical biologist) is concerned by this phase. It can be assisted by some software in order to validate normal or non- modified results. Medical staff are sometimes also required in order to explain pathology results to physicians. For a simple result given by phone or for a technical problem it's a medical technologist or medical lab scientist explaining it to a registered nurse. Departments in some countries are exclusively directed by a specialized pathologist, in others a consultant, medical or non-medical, may be the Head of Department. Clinical scientists have the right to interpret and discuss pathology results in their discipline in many countries, in Europe they are qualified to at least Masters level, may have a PhD and can have an exit qualification equivalent to medical staff (*e.g.*, FRCPath in the UK). In France, only medical staff (Pharm.D. and M.D. specialized in anatomical pathology or clinical biology) can discuss pathological results,

MEDICAL LABORATORY ACCREDITATION

Credibility of medical laboratories is paramount to the health and safety of the patients relying on the testing services provided by these labs. The international standard in use today for the accreditation of medical laboratories is ISO 15189 - Medical laboratories - Requirements for quality and competence. Accreditation is done by the Joint Commission, College of American Pathologists, AAB (American Association of Bioanalysts), and other state and federal agencies. CLIA 88 or the Clinical Laboratory Improvement Amendments also dictate testing and personnel. The accrediting body in Australia is NATA, all laboratories must be NATA accredited to receive payment from Medicare. In France, where accrediting body is COFRAC (COFRAC), in 2010, modification of legislation established ISO 15189 accreditation as an obligation for all clinical laboratories. In the United Arab Emirates, the Dubai Accreditation Department (DAC) is the accreditation body that is internationally recognised by the International Laboratory Accreditation Cooperation (ILAC) for many scopes including Medical Laboratories, Testing and Calibration Laboratories and Inspection Bodies.

In Hong Kong, the accrediting body is Hong Kong Accreditation Service (HKAS). On 16 February 2004, HKAS launched the medical testing accreditation programme. In Canada, laboratory accreditation is not mandatory, but is becoming more and more popular. Accreditation Canada (AC) is the national reference.

MENTAL STATUS EXAMINATION

The mental status examination or mental state examination (MSE) is an important part of the clinical assessment process in psychiatric practice. It is a structured way of observing and describing a patient's psychological functioning at a given point in time, under the domains of appearance, attitude, behaviour, mood, and affect, speech, thought process, thought content, perception, cognition, insight, and judgement. There are some minor variations in the subdivision of the MSE and the sequence and names of MSE domains. The purpose of the MSE is to obtain a comprehensive cross-sectional description of the patient's mental state, which, when combined with the biographical and historical information of the psychiatric history, allows the clinician to make an accurate diagnosis and formulation, which are required for coherent treatment planning. The data are collected through a combination of direct and indirect means: unstructured observation while obtaining the biographical and social information, focused questions about current symptoms, and formalised psychological tests. The MSE is not to be confused with the Mini-Mental State Examination (MMSE), which is a brief neuro-psychological screening test for dementia.

THEORETICAL FOUNDATIONS

The MSE derives from an approach to psychiatry known as descriptive psychopathology or descriptive phenomenology, which developed from the work of the philosopher and psychiatrist Karl Jaspers. From Jaspers' perspective it was assumed that the only way to comprehend a patient's experience is through his or her own description (through an approach of empathic and non-theoretical enquiry), as distinct from an interpretive or psychoanalytic approach which assumes the analyst might understand experiences or processes of which the patient is unaware, such as defence mechanisms or unconscious drives.

In practice, the MSE is a blend of empathic descriptive phenomenology and empirical clinical observation. It has been argued that the term *phenomenology* has become corrupted in clinical psychiatry: current usage, as a set of supposedly objective descriptions of a psychiatric patient (a synonym for signs and symptoms), is incompatible with the original meaning which was concerned with comprehending a patient's subjective experience.

APPLICATION

The mental status examination is a core skill of qualified (mental) health personnel. It is a key part of the initial psychiatric assessment in an out-patient or psychiatric hospital setting. It is a systematic collection of data based on observation of the patient's behaviour while the patient is in the clinician's view during the interview. The purpose is to obtain evidence of symptoms and signs of mental disorders, including danger to self and others, that are present at the time of the interview. Further, information on the patient's insight, judgement, and capacity for abstract reasoning is used to inform decisions about treatment strategy and the choice of an appropriate treatment setting.

It is carried out in the manner of an informal enquiry, using a combination of open and closed questions, supplemented by structured tests to assess cognition. The MSE can also be considered part of the comprehensive physical examination performed by physicians and nurses although it may be performed in a cursory and abbreviated way in non-mental-health settings. Information is usually recorded as free-form text using the standard headings, but brief MSE checklists are available for use in emergency situations, for example by paramedics or emergency department staff. The information obtained in the MSE is used, together with the biographical and social information of the psychiatric history, to generate a diagnosis, a psychiatric formulation and a treatment plan.

DOMAINS

Appearance

Clinicians assess the physical aspects such as the appearance of a patient, including apparent age, height, weight, and manner of dress and grooming. Colorful or bizarre clothing might suggest mania, while unkempt, dirty clothes might suggest schizophrenia or depression. If the patient appears much older than his or her chronological age this can suggest chronic poor self-care or ill-health. Clothing and accessories of a particular subculture, body modifications, or clothing not typical of the patient's gender, might give clues to personality. Observations of physical appearance might include the physical features of alcoholism or drug abuse, such as signs of malnutrition, nicotine stains, dental erosion, a rash around the mouth from inhalant abuse, or needle track marks from intravenous drug abuse. Observations can also include any odour which might suggest poor personal hygiene due to extreme self-neglect, or alcohol intoxication. Weight loss could also signify a depressive disorder, physical illness, anorexia nervosa or chronic anxiety.

Attitude

Attitude, also known as rapport, refers to the patient's approach to the interview process and the quality of information obtained during the assessment.

Behaviour

Abnormalities of behaviour, also called abnormalities of activity, include observations of specific abnormal movements, as well as more general observations of the patient's level of activity and arousal, and observations of the patient's eye contact and gait. Abnormal movements, for example choreiform, athetoid or choreoathetoid movements may indicate a neurological disorder. A tremor or dystonia may indicate a neurological condition or the side effects of antipsychotic medication. The patient may have tics (involuntary but quasi-purposeful movements or vocalizations) which may be a symptom of Tourette's syndrome. There are a range of abnormalities of movement which are typical of catatonia, such as echopraxia, catalepsy, waxy flexibility and paratonia (or *gegenhalten*). Stereotypies (repetitive purposeless movements such as rocking or head banging) or mannerisms (repetitive quasi-purposeful abnormal movements such as a gesture or abnormal gait) may be a feature of chronic schizophrenia or autism. More global behavioural abnormalities may be noted, such as an increase in arousal and movement (described as psychomotor agitation or hyperactivity) which might reflect mania or delirium. An inability to sit still might represent akathisia, a side effect of antipsychotic medication. Similarly, a global decrease in arousal and movement (described as psychomotor retardation, akinesia or stupor) might indicate depression or a medical condition such as Parkinson's disease, dementia or delirium. The examiner would also comment on eye movements (repeatedly glancing to one side can suggest that the patient is experiencing hallucinations), and the quality of eye contact (which can provide clues to the patient's emotional state). Lack of eye contact may suggest depression or autism.

Mood and Affect

The distinction between mood and affect in the MSE is subject to some disagreement. For example, Trzepacz and Baker (1993) describe affect as "the external and dynamic manifestations of a person's internal emotional state" and mood as "a person's predominant internal state at any one time", whereas Sims (1995) refers to affect as "differentiated specific feelings" and mood as "a more prolonged state or disposition". This article will use the Trzepacz and Baker (1993) definitions, with mood regarded as a current subjective state as described by the

patient, and affect as the examiner's inferences of the quality of the patient's emotional state based on objective observation. Mood is described using the patient's own words, and can also be described in summary terms such as neutral, euthymic, dysphoric, euphoric, angry, anxious or apathetic. Alexithymic individuals may be unable to describe their subjective mood state. An individual who is unable to experience any pleasure may be suffering from anhedonia. Affect is described by labelling the apparent emotion conveyed by the person's nonverbal behaviour (anxious, sad etc.), and also by using the parameters of appropriateness, intensity, range, reactivity and mobility. Affect may be described as appropriate or inappropriate to the current situation, and as congruent or incongruent with their thought content. For example, someone who shows a bland affect when describing a very distressing experience would be described as showing incongruent affect, which might suggest schizophrenia. The intensity of the affect may be described as normal, blunted affect, exaggerated, flat, heightened or overly dramatic. A flat or blunted affect is associated with schizophrenia, depression or post-traumatic stress disorder; heightened affect might suggest mania, and an overly dramatic or exaggerated affect might suggest certain personality disorders. Mobility refers to the extent to which affect changes during the interview: the affect may be described as fixed, mobile, immobile, constricted/restricted or labile. The person may show a full range of affect, in other words a wide range of emotional expression during the assessment, or may be described as having restricted affect. The affect may also be described as reactive, in other words changing flexibly and appropriately with the flow of conversation, or as unreactive. A bland lack of concern for one's disability may be described as showing *la belle indifférence*, a feature of conversion disorder, which is historically termed "hysteria" in older texts.

Speech

The patient's speech is assessed by observing the patient's spontaneous speech, and also by using structured tests of specific language functions. This heading is concerned with the production of speech rather than the *content* of speech, which is addressed under thought form and thought content. When observing the patient's spontaneous speech, the interviewer will note and comment on paralinguistic features such as the loudness, rhythm, prosody, intonation, pitch, phonation, articulation, quantity, rate, spontaneity and latency of speech. A structured assessment of speech includes an assessment of expressive language by asking the patient to name objects, repeat short sentences, or produce as many words as possible from a certain category in a set time. Simple language tests form part of the mini-mental state examination. In practice, the structured assessment of receptive and expressive language is often reported under Cognition. Language assessment will allow the recognition of medical conditions presenting with aphonia or dysarthria, neurological conditions such as stroke or dementia presenting with aphasia, and specific language disorders such as stuttering, cluttering or mutism. People with autism or Asperger syndrome may have abnormalities in paralinguistic and pragmatic aspects of their speech. Echolalia (repetition of another person's words) and palilalia (repetition of the subject's own words) can be heard with patients with autism, schizophrenia or Alzheimer's disease. A person with schizophrenia might use neologisms, which are made-up words which have a specific meaning to the person using them. Speech assessment also contributes to assessment of mood, for example people with mania or anxiety may have rapid, loud and pressured speech; on the other hand depressed patients will typically have a prolonged speech latency and speak in a slow, quiet and hesitant manner.

Thought Process

Thought process in the MSE refers to the quantity, tempo (rate of flow) and form (or logical coherence) of thought. Thought process cannot be directly observed but can only be described by the patient, or inferred from a patient's speech. Regarding the tempo of thought, some people may experience flight of ideas, when their thoughts are so rapid that their speech seems incoherent, although a careful observer can discern a chain of

poetic associations in the patient's speech. Alternatively an individual may be described as having retarded or inhibited thinking, in which thoughts seem to progress slowly with few associations. Poverty of thought is a global reduction in the quantity of thought and thought perseveration refers to a pattern where a person keeps returning to the same limited set of ideas. A pattern of interruption or disorganization of thought processes is broadly referred to as formal thought disorder, and might be described more specifically as thought blocking, fusion, loosening of associations, tangential thinking, derailment of thought, or knight's move thinking. Thought may be described as circumstantial when a patient includes a great deal of irrelevant detail and makes frequent diversions, but remains focused on the broad topic. Flight of ideas is typical of mania. Conversely, patients with depression may have retarded or inhibited thinking. Poverty of thought is one of the negative symptoms of schizophrenia, and might also be a feature of severe depression or dementia. A patient with dementia might also experience thought perseveration. Formal thought disorder is a common feature of schizophrenia. Circumstantial thinking might be observed in anxiety disorders or certain kinds of personality disorders.

Thought Content

A description of thought content would describe a patient's delusions, overvalued ideas, obsessions, phobias and preoccupations. Abnormalities of thought content are established by exploring individuals' thoughts in an open-ended conversational manner with regard to their intensity, salience, the emotions associated with the thoughts, the extent to which the thoughts are experienced as one's own and under one's control, and the degree of belief or conviction associated with the thoughts. A delusion can be defined as "a false, unshakeable idea or belief which is out of keeping with the patient's educational, cultural and social background... held with extraordinary conviction and subjective certainty", and is a core feature of psychotic disorders. The patient's delusions may be described as persecutory or paranoid delusions, delusions of reference, grandiose delusions, erotomanic delusions, delusional jealousy or delusional misidentification. Delusions may be described as mood-congruent (the delusional content in keeping with the mood), typical of manic or depressive psychoses, or mood-incongruent (delusional content not in keeping with the mood) which are more typical of schizophrenia. Delusions of control, or passivity experiences (in which the individual has the experience of the mind or body being under the influence or control of some kind of external force or agency), are typical of schizophrenia. Examples of this include experiences of thought withdrawal, thought insertion, thought broadcasting, and somatic passivity. Schneiderian first rank symptoms are a set of delusions and hallucinations which have been said to be highly suggestive of a diagnosis of schizophrenia. Delusions of guilt, delusions of poverty, and nihilistic delusions (belief that one has no mind or is already dead) are typical of depressive psychoses. An overvalued idea is a false belief that is held with conviction but not with delusional intensity. Hypochondriasis is an overvalued idea that one is suffering from an illness, dysmorphophobia is an overvalued idea that a part of one's body is abnormal, and people with anorexia nervosa may have an overvalued idea of being thin.

An obsession is an "undesired, unpleasant, intrusive thought that cannot be suppressed through the patient's volition", but unlike passivity experiences described above, they are not experienced as imposed from outside the patient's mind. Obsessions are typically intrusive thoughts of violence, injury, dirt or sex, or obsessive ruminations on intellectual themes. A person can also describe obsessional doubt, with intrusive worries about whether they have made the wrong decision, or forgotten to do something, for example turn off the gas or lock the house. In obsessive-compulsive disorder, the individual experiences obsessions with or without compulsions (a sense of having to carry out certain ritualized and senseless actions against their wishes). A phobia is "a dread of an object or situation that does not in reality pose any threat", and is distinct from a delusion in that the patient is aware that the fear is irrational. A phobia is usually highly specific to certain situations and will usually be reported by the patient rather than being observed by the clinician in the assessment interview. Preoccupations are thoughts which are not fixed, false or intrusive, but have an undue prominence in the

person's mind. Clinically significant preoccupations would include thoughts of suicide, homicidal thoughts, suspicious or fearful beliefs associated with certain personality disorders, depressive beliefs (for example that one is unloved or a failure), or the cognitive distortions of anxiety and depression. The MSE contributes to clinical risk assessment by including a thorough exploration of any suicidal or hostile thought content. Assessment of suicide risk includes detailed questioning about the nature of the person's suicidal thoughts, belief about death, reasons for living, and whether the person has made any specific plans to end his or her life.

Perceptions

A perception in this context is any sensory experience, and the three broad types of perceptual disturbance are hallucinations, pseudohallucinations and illusions. A hallucination is defined as a sensory perception in the absence of any external stimulus, and is experienced in external or objective space (*i.e.* experienced by the subject as real). An illusion is defined as a false sensory perception in the presence of an external stimulus, in other words a distortion of a sensory experience, and may be recognized as such by the subject. A pseudohallucination is experienced in internal or subjective space (for example as "voices in my head") and is regarded as akin to fantasy. Other sensory abnormalities include a distortion of the patient's sense of time, for example *déjà vu*, or a distortion of the sense of self (depersonalization) or sense of reality (derealization). Hallucinations can occur in any of the five senses, although auditory and visual hallucinations are encountered more frequently than tactile (touch), olfactory (smell) or gustatory (taste) hallucinations. Auditory hallucinations are typical of psychoses: third-person hallucinations (*i.e.* voices talking about the patient) and hearing one's thoughts spoken aloud (*gedankenlautwerden* or *écho de la pensée*) are among the Schneiderian first rank symptoms indicative of schizophrenia, whereas second-person hallucinations (voices talking to the patient) threatening or insulting or telling them to commit suicide, may be a feature of psychotic depression or schizophrenia. Visual hallucinations are generally suggestive of organic conditions such as epilepsy, drug intoxication or drug withdrawal. Many of the visual effects of hallucinogenic drugs are more correctly described as visual illusions or visual pseudohallucinations, as they are distortions of sensory experiences, and are not experienced as existing in objective reality. Auditory pseudohallucinations are suggestive of dissociative disorders. *Déjà vu*, derealization and depersonalization are associated with temporal lobe epilepsy and dissociative disorders.

Cognition

This section of the MSE covers the patient's level of alertness, orientation, attention, memory, visuospatial functioning, language functions and executive functions. Unlike other sections of the MSE, use is made of structured tests in addition to unstructured observation. Alertness is a global observation of level of consciousness *i.e.* awareness of, and responsiveness to the environment, and this might be described as alert, clouded, drowsy, or stuporous. Orientation is assessed by asking the patient where he or she is (for example what building, town and state) and what time it is (time, day, date).

Attention and concentration are assessed by the serial sevens test (or alternatively by spelling a five-letter word backwards), and by testing digit span. Memory is assessed in terms of immediate registration (repeating a set of words), short-term memory (recalling the set of words after an interval, or recalling a short paragraph), and long-term memory (recollection of well known historical or geographical facts). Visuospatial functioning can be assessed by the ability to copy a diagram, draw a clock face, or draw a map of the consulting room. Language is assessed through the ability to name objects, repeat phrases, and by observing the individual's spontaneous speech and response to instructions. Executive functioning can be screened for by asking the "similarities" questions ("what do x and y have in common?") and by means of a verbal fluency task (*e.g.* "list as many words as you can starting with the letter F, in one minute"). The mini-mental state examination is a simple structured cognitive assessment which is in widespread use as a component of the MSE.

Mild impairment of attention and concentration may occur in any mental illness where people are anxious and distractible (including psychotic states), but more extensive cognitive abnormalities are likely to indicate a gross disturbance of brain functioning such as delirium, dementia or intoxication. Specific language abnormalities may be associated with pathology in Wernicke's area or Broca's area of the brain. In Korsakoff's syndrome there is dramatic memory impairment with relative preservation of other cognitive functions. Visuospatial or constructional abnormalities here may be associated with parietal lobe pathology, and abnormalities in executive functioning tests may indicate frontal lobe pathology. This kind of brief cognitive testing is regarded as a screening process only, and any abnormalities are more carefully assessed using formal neuropsychological testing.

The MSE may include a brief neuropsychiatric examination in some situations. Frontal lobe pathology is suggested if the person cannot repetitively execute a motor sequence (*e.g.* "paper-scissors-stone"). The posterior columns are assessed by the person's ability to feel the vibrations of a tuning fork on the wrists and ankles. The parietal lobe can be assessed by the person's ability to identify objects by touch alone and with eyes closed. A cerebellar disorder may be present if the person cannot stand with arms extended, feet touching and eyes closed without swaying (Romberg's sign); if there is a tremor when the person reaches for an object; or if he or she is unable to touch a fixed point, close the eyes and touch the same point again. Pathology in the basal ganglia may be indicated by rigidity and resistance to movement of the limbs, and by the presence of characteristic involuntary movements. A lesion in the posterior fossa can be detected by asking the patient to roll his or her eyes upwards (Parinaud's syndrome). Focal neurological signs such as these might reflect the effects of some prescribed psychiatric medications, chronic drug or alcohol use, head injuries, tumors or other brain disorders.

Insight

The person's understanding of his or her mental illness is evaluated by exploring his or her explanatory account of the problem, and understanding of the treatment options. In this context, insight can be said to have three components: recognition that one has a mental illness, compliance with treatment, and the ability to re-label unusual mental events (such as delusions and hallucinations) as pathological. As insight is on a continuum, the clinician should not describe it as simply present or absent, but should report the patient's explanatory account descriptively. Impaired insight is characteristic of psychosis and dementia, and is an important consideration in treatment planning and in assessing the capacity to consent to treatment.

Judgement

Judgement refers to the patient's capacity to make sound, reasoned and responsible decisions. Traditionally, the MSE included the use of standard hypothetical questions such as "what would you do if you found a stamped, addressed envelope lying in the street?"; however contemporary practice is to inquire about how the patient has responded or would respond to real-life challenges and contingencies. Assessment would take into account the individual's executive system capacity in terms of impulsiveness, social cognition, self-awareness and planning ability. Impaired judgement is not specific to any diagnosis but may be a prominent feature of disorders affecting the frontal lobe of the brain. If a person's judgement is impaired due to mental illness, there might be implications for the person's safety or the safety of others.

CULTURAL CONSIDERATIONS

There are potential problems when the MSE is applied in a cross-cultural context, when the clinician and patient are from different cultural backgrounds. For example, the patient's culture might have different norms for appearance, behaviour and display of emotions. Culturally normative spiritual and religious beliefs need to

be distinguished from delusions and hallucinations - these may seem similar to one who does not understand that they have different roots. Cognitive assessment must also take the patient's language and educational background into account. Clinician's racial bias is another potential confounder.

CHILDREN

There are particular challenges in carrying out an MSE with young children and others with limited language such as people with intellectual impairment. The examiner would explore and clarify the individual's use of words to describe mood, thought content or perceptions, as words may be used idiosyncratically with a different meaning from that assumed by the examiner. In this group, tools such as play materials, puppets, art materials or diagrams (for instance with multiple choices of facial expressions depicting emotions) may be used to facilitate recall and explanation of experiences.

MICROVESICLES

Microvesicles (circulating microvesicles, or microparticles) are a type of extracellular vesicle, between 50 and 1,000 nanometers (nm) in diameter, found in many types of body fluids as well as the interstitial space between cells. Microvesicles are membrane-bound vesicles containing phospholipids, ranging from 100 nm to 1000 nm shed from almost all cell types. Not to be confused with smaller intracellularly generated extracellular vesicles known as exosomes. Microvesicles play a role in intercellular communication and can transport mRNA, miRNA, and proteins between cells. Microvesicles have been implicated in the process of a remarkable anti-tumor reversal effect in cancer, tumor immune suppression, metastasis, tumor-stroma interactions and angiogenesis along with having a primary role in tissue regeneration. They originate directly from the plasma membrane of the cell and reflect the antigenic content of the cells from which they originate. They remove misfolded proteins, cytotoxic agents and metabolic waste from the cell. Though initially dismissed as cellular debris, microvesicles have a role in cell signaling and the process of molecular communication between cells, and are released by a number of cell types. Although a consistent and precise definition is lacking, microvesicles are generally considered to be a heterogeneous population of exosomes (<100 nm) and shed microvesicles (100-1000 nm), which are similar but have distinct mechanisms of formation. Through these mechanisms, microvesicles are released into the extracellular space and interact with specific target cells, delivering bioactive molecules. Changes in microvesicle levels are implicated in a variety of diseases, including cancer. These changes can be used as biomarkers in a variety of diagnostic assays.

FORMATION AND CONTENTS

Different cells can release microvesicles from the plasma membrane. Sources of microvesicles include megakaryocytes, blood platelets, monocytes, neutrophils, tumor cells and placenta. Platelets play an important role in maintaining hemostasis: they promote thrombus growth, and thus they prevent loss of blood. Moreover, they enhance immune response, since they express the molecule CD154 (CD40L). Platelets are activated by inflammation, infection, or injury, and after their activation microvesicles containing CD154 are released from platelets. CD154 is a crucial molecule in the development of T cell-dependent humoral immune response. CD154 knockout mice are incapable of producing IgG, IgE, or IgA as a response to antigens. Microvesicles can also transfer prions and molecules CD41 and CXCR4.

Endothelial Microparticles

Endothelial microparticles are small vesicles that are released from endothelial cells and can be found circulating in the blood. The microparticle consists of a plasma membrane surrounding a small amount of

cytosol. The membrane of the endothelial microparticle contains receptors and other cell surface molecules which enable the identification of the endothelial origin of the microparticle, and allow it to be distinguished from microparticles from other cells, such as platelets. Although circulating endothelial microparticles can be found in the blood of normal individuals, increased numbers of circulating endothelial microparticles have been identified in individuals with certain diseases, including hypertension and cardiovascular disorders, and pre-eclampsia and various forms of vasculitis. The endothelial microparticles in some of these disease states have been shown to have arrays of cell surface molecules reflecting a state of endothelial dysfunction. Therefore, endothelial microparticles may be useful as an indicator or index of the functional state of the endothelium in disease, and may potentially play key roles in the pathogenesis of certain diseases, including rheumatoid arthritis.

Process of Formation

Microvesicles and exosomes are formed and released by two slightly different mechanisms. These processes result in the release of intercellular signaling vesicles. Microvesicles are small, plasma membrane-derived particles that are released into the extracellular environment by the outward budding and fission of the plasma membrane. This budding process involves multiple signaling pathways including the elevation of intracellular calcium and reorganization of the cell's structural scaffolding. The formation and release of microvesicles involve contractile machinery that draws opposing membranes together before pinching off the membrane connection and launching the vesicle into the extracellular space. Microvesicle budding takes place at unique locations on the cell membrane that are enriched with specific lipids and proteins reflecting their cellular origin. At these locations, proteins, lipids, and nucleic acids are selectively incorporated into microvesicles and released into the surrounding environment.

Exosomes are membrane-covered vesicles, formed intracellularly are considered to be smaller than 100 nm. In contrast to microvesicles, which are formed through a process of membrane budding, or exocytosis, exosomes are initially formed by endocytosis. Exosomes are formed by invagination within a cell to create an intracellular vesicle called an endosome, or an endocytic vesicle. In general, exosomes are formed by segregating the cargo (*e.g.*, lipids, proteins, and nucleic acids) within the endosome. Once formed, the endosome combines with a structure known as a multivesicular body (MVB). The MVB containing segregated endosomes ultimately fuses with the plasma membrane, resulting in exocytosis of the exosomes. Once formed, both microvesicles and exosomes (collectively called microvesicles) circulate in the extracellular space near the site of release, where they can be taken up by other cells or gradually deteriorate. In addition, some vesicles migrate significant distances by diffusion, ultimately appearing in biological fluids such as cerebrospinal fluid, blood, and urine.

Mechanism of Shedding

There are three mechanisms which lead to release of vesicles into the extracellular space. First of these mechanisms is exocytosis from multivesicular bodies and the formation of exosomes. Another mechanism is budding of microvesicles directly from a plasma membrane. And the last one is cell death leading to apoptotic blebbing. These are all energy-requiring processes. Under physiologic conditions, the plasma membrane of cells has an asymmetric distribution of phospholipids. aminophospholipids, phosphatidylserine, and phosphatidylethanolamine are specifically sequestered in the inner leaflet of the membrane. The transbilayer lipid distribution is under the control of three phospholipidic pumps: an inward-directed pump, or flippase; an outward-directed pump, or floppase; and a lipid scramblase, responsible for non-specific redistribution of lipids across the membrane. After cell stimulation, including apoptosis, a subsequent cytosolic Ca increase promotes the loss of phospholipid asymmetry of the plasma membrane, subsequent phosphatidylserine exposure, and a transient phospholipidic imbalance between the external leaflet at the expense of the inner leaflet, leading to budding of the plasma membrane and microvesicle release.

Molecular Contents

The lipid and protein content of microvesicles has been analyzed using various biochemical techniques. Microvesicles display a spectrum of enclosed molecules enclosed within the vesicles and their plasma membranes. Both the membrane molecular pattern and the internal contents of the vesicle depend on the cellular origin and the molecular processes triggering their formation. Because microvesicles are not intact cells, they do not contain mitochondria, Golgi, endoplasmic reticulum, or a nucleus with its associated DNA. Microvesicle membranes consist mainly of membrane lipids and membrane proteins. Regardless of their cell type of origin, nearly all microvesicles contain proteins involved in membrane transport and fusion. They are surrounded by a phospholipid bilayer composed of several different lipid molecules. The protein content of each microvesicle reflects the origin of the cell from which it was released. For example, those released from antigen-presenting cells (APCs), such as B cells and dendritic cells, are enriched in proteins necessary for adaptive immunity, while microvesicles released from tumors contain proapoptotic molecules and oncogenic receptors (*e.g.* EGFR). In addition to the proteins specific to the cell type of origin, some proteins are common to most microvesicles. For example, nearly all contain the cytoplasmic proteins tubulin, actin and actin-binding proteins, as well as many proteins involved in signal transduction, cell structure and motility, and transcription. Most microvesicles contain the so-called “heat-shock proteins” hsp70 and hsp90, which can facilitate interactions with cells of the immune system. Finally, tetraspanin proteins, including CD9, CD37, CD63 and CD81 are one of the most abundant protein families found in microvesicle membranes. Many of these proteins may be involved in the sorting and selection of specific cargos to be loaded into the lumen of the microvesicle or its membrane. Other than lipids and proteins, microvesicles are enriched with nucleic acids (*e.g.*, messenger RNA (mRNA) and microRNA (miRNA)). The identification of RNA molecules in microvesicles supports the hypothesis that they are a biological vehicle for the transfer of nucleic acids and subsequently modulate the target cell’s protein synthesis.

Messenger RNA transported from one cell to another through microvesicles can be translated into proteins, conferring new function to the target cell. The discovery that microvesicles may shuttle specific mRNA and miRNA suggests that this may be a new mechanism of genetic exchange between cells. Exosomes produced by cells exposed to oxidative stress can mediate protective signals, reducing oxidative stress in recipient cells, a process which is proposed to depend on exosomal RNA transfer. These RNAs are specifically targeted to microvesicles, in some cases containing detectable levels of RNA that is not found in significant amounts in the donor cell.

Because the specific proteins, mRNAs, and miRNAs in microvesicles are highly variable, it is likely that these molecules are specifically packaged into vesicles using an active sorting mechanism. At this point, it is unclear exactly which mechanisms are involved in packaging soluble proteins and nucleic acids into microvesicles.

ROLE ON TARGET CELLS

Once released from their cell of origin, microvesicles interact specifically with cells they recognize by binding to cell-type specific, membrane-bound receptors. Because microvesicles contain a variety of surface molecules, they provide a mechanism for engaging different cell receptors and exchanging material between cells.

This interaction ultimately leads to fusion with the target cell and release of the vesicles’ components, thereby transferring bioactive molecules, lipids, genetic material, and proteins. The transfer of microvesicle components includes specific mRNAs and proteins, contributing to the proteomic properties of target cells. microvesicles can also transfer miRNAs that are known to regulate gene expression by altering mRNA turnover.

Mechanisms of Signaling

Degradation

In some cases, the degradation of microvesicles is necessary for the release of signaling molecules. During microvesicle production, the cell can concentrate and sort the signaling molecules which are released into the extracellular space upon microvesicle degradation. Dendritic cells, macrophage and microglia derived microvesicles contain proinflammatory cytokines and neurons and endothelial cells release growth factors using this mechanism of release.

Fusion

Proteins on the surface of the microvesicle will interact with specific molecules, such as integrin, on the surface of its target cell. Upon binding, the microvesicle can fuse with the plasma membrane. This results in the delivery of nucleotides and soluble proteins into the cytosol of the target cell as well as the integration of lipids and membrane proteins into its plasma membrane.

Internalization

Microvesicle can be endocytosed upon binding to their targets, allowing for additional steps of regulation by the target cell. The microvesicle may fuse, integrating lipids and membrane proteins into the endosome while releasing its contents into the cytoplasm. Alternatively, the endosome may mature into a lysosome causing the degradation of the microvesicle and its contents, in which case the signal is ignored.

Transcytosis

After internalization of microvesicle via endocytosis, the endosome may move across the cell and fuse with the plasma membrane, a process called transcytosis. This results in the ejection of the microvesicle back into the extracellular space or may result in the transportation of the microvesicle into a neighboring cell. This mechanism might explain the ability of microvesicle to cross biological barriers, such as the blood brain barrier, by moving from cell to cell.

Contact Dependent Signaling

In this form of signaling, the microvesicle does not fuse with the plasma membrane or engulfed by the target cell. Similar to the other mechanisms of signaling, the microvesicle has molecules on its surface that will interact specifically with its target cell. There are additional surface molecules, however, that can interact with receptor molecules which will interact with various signaling pathways. This mechanism of action can be used in processes such as antigen presentation, where MHC molecules on the surface of microvesicle can stimulate an immune response. Alternatively, there may be molecules on microvesicle surfaces that can recruit other proteins to form extracellular protein complexes that may be involved in signaling to the target cell.

RELEVANCE IN DISEASE

Cancer

Promoting Aggressive Tumor Phenotypes

The oncogenic receptor ECGFvIII, which is located in a specific type of aggressive glioma tumor, can be transferred to a non-aggressive population of tumor cells via microvesicles. After the oncogenic protein is

transferred, the recipient cells become transformed and show characteristic changes in the expression levels of target genes. It is possible that transfer of other mutant oncogenes, such as HER2, may be a general mechanism by which malignant cells cause cancer growth at distant sites. microvesicles from non-cancer cells can signal to cancer cells to become more aggressive. Upon exposure to microvesicles from tumor-associated macrophages, breast cancer cells become more invasive *in vitro*.

Promoting Angiogenesis

Angiogenesis, which is essential for tumor survival and growth, occurs when endothelial cells proliferate to create a matrix of blood vessels that infiltrate the tumor, supplying the nutrients and oxygen necessary for tumor growth. A number of reports have demonstrated that tumor-associated microvesicles release proangiogenic factors that promote endothelial cell proliferation, angiogenesis, and tumor growth. microvesicles shed by tumor cells and taken up by endothelial cells also facilitate angiogenic effects by transferring specific mRNAs and miRNAs.

Involvement in Multidrug Resistance

When anticancer drugs such as doxorubicin accumulate in microvesicles, the drug's cellular levels decrease. This can ultimately contribute to the process of drug resistance. Similar processes have been demonstrated in microvesicles released from cisplatin-insensitive cancer cells. Vesicles from these tumors contained nearly three times more cisplatin than those released from cisplatin-sensitive cells. For example, tumor cells can accumulate drugs into microvesicles. Subsequently, the drug-containing microvesicles are released from the cell into the extracellular environment, thereby mediating resistance to chemotherapeutic agents and resulting in significantly increased tumor growth, survival, and metastasis.

Interference with Antitumor Immunity

microvesicles from various tumor types can express specific cell-surface molecules (*e.g.* FasL or CD95) that induce T-cell apoptosis and reduce the effectiveness of other immune cells. microvesicles released from lymphoblastoma cells express the immune-suppressing protein latent membrane protein-1 (LMP-1), which inhibits T-cell proliferation and prevents the removal of circulating tumor cells (CTCs). As a consequence, tumor cells can turn off T-cell responses or eliminate the antitumor immune cells altogether by releasing microvesicles.

Impact on Tumor Metastasis

Degradation of the extracellular matrix is a critical step in promoting tumor growth and metastasis. Tumor-derived microvesicles often carry protein-degrading enzymes, including matrix metalloproteinase 2 (MMP-2), MMP-9, and urokinase-type plasminogen activator (uPA). By releasing these proteases, tumor cells can degrade the extracellular matrix and invade surrounding tissues. Likewise, inhibiting MMP-2, MMP-9, and uPA prevents microvesicles from facilitating tumor metastasis. Matrix digestion can also facilitate angiogenesis, which is important for tumor growth and is induced by the horizontal transfer of RNAs from microvesicles.

Other Disease States

The release of microvesicles has been shown from endothelial cells, vascular smooth muscle cells, platelets, white blood cells (*e.g.* leukocytes and lymphocytes), and red blood cells. Although some of these microvesicle populations occur in the blood of healthy individuals and patients, there are obvious changes in number, cellular

origin, and composition in various disease states. It has become clear that microvesicles play important roles in regulating the cellular processes that lead to disease pathogenesis. Moreover, because microvesicles are released following apoptosis or cell activation, they have the potential to induce or amplify disease processes. Some of the inflammatory and pathological conditions that microvesicles are involved in include cardiovascular disease, hypertension, neurodegenerative disorders, diabetes, and rheumatic diseases.

Vascular Diseases

Circulating microvesicles isolated from cardiac surgery patients were found to be thrombogenic in both in vitro assays and in rats. Microvesicles isolated from healthy individuals did not have the same effects and may actually have a role in reducing clotting. Tissue factor, an initiator of coagulation, is found in high levels within microvesicles, indicating their role in clotting. Additionally, microvesicles can induce clotting by binding to clotting factors or by inducing the expression of clotting factors in other cells. Microvesicles and tissue factor are associated with diabetic vasculopathy in a mechanism affected by hyperglycemia in diabetic patients. Renal mesangial cells exposed to high glucose media release microvesicles containing tissue factor, having an angiogenic effect on endothelial cells. Atherosclerosis has also been linked with circulating microvesicles originating from platelets and macrophages. These microvesicles are found in high levels within atherosclerotic plaques, and their presence results in communication with clotting machinery that exacerbates the condition.

Inflammation

Microvesicles contain cytokines that can induce inflammation via numerous different pathways. These cells will then release more microvesicles, which have an additive effect. This can call neutrophils and leukocytes to the area, resulting in the aggregation of cells. However, microvesicles also seem to be involved in a normal physiological response to disease, as there are increased levels of microvesicles that result from pathology.

Neurological Disorders

Microvesicles seem to be involved in a number of neurological diseases. Since they are involved in numerous vascular diseases and inflammation, strokes and multiple sclerosis seem to be other diseases for which microvesicles are involved. Circulating microvesicles seem to have an increased level of phosphorylated tau proteins during early stage Alzheimer's disease. Similarly, increased levels of CD133 are an indicator of epilepsy.

CLINICAL APPLICATIONS

Detection of Cancer

Tumor-associated microvesicles are abundant in the blood, urine, and other body fluids of patients with cancer, and are likely involved in tumor progression. They offer a unique opportunity to noninvasively access the wealth of biological information related to their cells of origin. The quantity and molecular composition of microvesicles released from malignant cells varies considerably compared with those released from normal cells. Thus, the concentration of plasma microvesicles with molecular markers indicative of the disease state may be used as an informative blood-based biosignature for cancer. Microvesicles express many membrane-bound proteins, some of which can be used as tumor biomarkers. Several tumor markers accessible as proteins in blood or urine have been used to screen and diagnose various types of cancer. In general, tumor markers are produced either by the tumor itself or by the body in response to the presence of cancer or some inflammatory conditions. If a tumor marker level is higher than normal, the patient is examined more closely to look for

cancer or other conditions. For example, CA19-9, CA-125, and CEA have been used to help diagnose pancreatic, ovarian, and gastrointestinal malignancies, respectively. However, although they have proven clinical utility, none of these tumor markers are highly sensitive or specific. Clinical research data suggest that tumor-specific markers exposed on microvesicles are useful as a clinical tool to diagnose and monitor disease. Research is also ongoing to determine if tumor-specific markers exposed on microvesicles are predictive for therapeutic response. Evidence produced by independent research groups has demonstrated that microvesicles from the cells of healthy tissues, or selected miRNAs from these microvesicles, can be employed to reverse many tumors in pre-clinical cancer models, and may be used in combination with chemotherapy. Conversely, microvesicles processed from a tumor cell are involved in the transport of cancer proteins and in delivering microRNA to the surrounding healthy tissue. It leads to a change of healthy cell phenotype and creates a tumor-friendly environment. Microvesicles play an important role in tumor angiogenesis and in the degradation of matrix due to the presence of metalloproteases, which facilitate metastasis. They are also involved in intensification of the function of regulatory T-lymphocytes and in the induction of apoptosis of cytotoxic T-lymphocytes, because microvesicles released from a tumor cell contain Fas ligand and TRAIL. They prevent differentiation of monocytes to dendritic cells. Tumor microvesicles also carry tumor antigen, so they can be an instrument for developing tumor vaccines. Circulating miRNA and segments of DNA in all body fluids can be potential markers for tumor diagnostics.

Microvesicles and Rheumatoid Arthritis

Rheumatoid arthritis is a chronic systemic autoimmune disease characterized by inflammation of joints. In the early stage there are abundant Th17 cells producing proinflammatory cytokines IL-17A, IL-17F, TNF, IL-21, and IL-22 in the synovial fluid. regulatory T-lymphocytes have a limited capability to control these cells. In the late stage, the extent of inflammation correlates with numbers of activated macrophages that contribute to joint inflammation and bone and cartilage destruction, because they have the ability to transform themselves into osteoclasts that destroy bone tissue. Synthesis of reactive oxygen species, proteases, and prostaglandins by neutrophils is increased. Activation of platelets via collagen receptor GPVI stimulates the release of microvesicles from platelet cytoplasmic membranes. These microparticles are detectable at a high level in synovial fluid, and they promote joint inflammation by transporting proinflammatory cytokine IL-1.

Biological Markers for Disease

In addition to detecting cancer, it is possible to use microvesicles as biological markers to give prognoses for various diseases. Many types of neurological diseases are associated with increased level of specific types of circulating microvesicles. For example, elevated levels of phosphorylated tau proteins can be used to diagnose patients in early stages of Alzheimer's. Additionally, it is possible to detect increased levels of CD133 in microvesicles of patients with epilepsy.

Mechanism for Drug Delivery

Circulating microvesicles may be useful for the delivery of drugs to very specific targets. Using electroporation or centrifugation to insert drugs into microvesicles targeting specific cells, it is possible to target the drug very efficiently. This targeting can help by reducing necessary doses as well as prevent off-target side effects. They can target anti-inflammatory drugs to specific tissues. Additionally, circulating microvesicles can bypass the blood–brain barrier and deliver their cargo to neurons while not having an effect on muscle cells. The blood–brain barrier is typically a difficult obstacle to overcome when designing drugs, and microvesicles may be a means of overcoming it. Current research is looking into efficiently creating microvesicles synthetically, or isolating them from patient or engineered cell lines.

SUTTON'S LAW

Sutton's law states that when diagnosing, one should first consider the obvious. It suggests that one should first conduct those tests which could confirm (or rule out) the most likely diagnosis. It is taught in medical schools to suggest to medical students that they might best order tests in that sequence which is most likely to result in a quick diagnosis, hence treatment, while minimizing unnecessary costs. It is also applied in pharmacology, when choosing a drug to treat a specific disease you want the drug to reach the disease. It is applicable to any process of diagnosis, *e.g.* debugging computer programmes. Computer-aided diagnosis provides a statistical and quantitative approach. A more thorough analysis will consider the false positive rate of the test and the possibility that a less likely diagnosis might have more serious consequences. A competing principle is the idea of performing simple tests before more complex and expensive tests, moving from bedside tests to blood results and simple imaging such as ultrasound and then more complex such as MRI then specialty imaging. The law can also be applied in prioritizing tests when resources are limited, so a test for a treatable condition should be performed before an equally probable but less treatable condition. The law is named after the bank robber Willie Sutton, who reputedly replied to a reporter's inquiry as to why he robbed banks by saying "because that's where the money is." In Sutton's 1976 book *Where the Money Was*, Sutton denies having said this. A similar idea is contained in the physician's adage, "When you hear hoofbeats, think horses, not zebras."

SURGICAL SIEVE

The surgical sieve is a thought process in medicine. It is a typical example of how to organise a structured examination answer for medical students and physicians when they are challenged with a question. It is also a way of constructing answers to questions from patients and their relatives in a logical manner, and structuring articles and reference texts in medicine. Some textbooks put emphasis on using the surgical sieve as a basic structure of diagnosis and management of illnesses.

Although there are several versions around the world with slight variations, the surgical sieve usually consist of the following types of process in the human body in any particular order:

- Congenital
- Acquired
 - (a) Vascular
 - (b) Infective
 - (c) Traumatic
 - (d) Autoimmune
 - (e) Metabolic
 - (f) Inflammatory
 - (g) Neurological
 - (h) Neoplastic
 - (i) Degenerative
 - (j) Environmental
 - (k) Unknown

A more extensive, and perhaps more concise mechanism of employing the surgical sieve is using the mnemonic MEDIC HAT PINE:

- Metabolic (conditions relating to metabolism, biochemistry, and the like)
- Endocrinological (conditions relating to the various secretory systems within the body)
- Degenerative (conditions relating to age-related destruction of tissue, or stress-related destruction of tissue)
- Inflammatory/Infective (conditions that primarily present in a way that involves the profane activation of the immune system)

- Congenital (conditions present at birth)
- Haematological (conditions relating to the blood system, in one way or another)
- Autoimmune (conditions relating to the inappropriate activation of the immune system, in one of many ways)
- Traumatic (conditions relating to a physical response between two or more objects)
- Psychological/Neurological (conditions relating to the nervous system, in one way or another – whether that be the CNS or the PNS)
- Idiopathic/iatrogenic (conditions without a known cause, or without a known cause outside of medical intervention)
- Neoplastic (conditions relating to cancers)
- Environmental (conditions relating to exposures, and dose-response relationships thereof)

EXAMPLES

What are the causes of an acute confusional state in a patient?

1. Treatment induced: polypharmacy, sedatives, analgesics, steroids, drug withdrawal
2. Vascular: stroke, TIA, vascular dementia
3. Inflammatory: infection, systemic inflammatory response syndrome
4. Traumatic: head injury, Intracranial hemorrhage, shock
5. Autoimmune: thyroid disease
6. Metabolic: electrolyte imbalance, DKA, hypoglycaemia, SIADH
7. Infective: sepsis, local infection
8. Neoplastic: brain tumour, carcinomatosis
9. Degenerative: Alzheimer's Disease, dementia

What are the causes of splenomegaly?

1. Idiopathic: Idiopathic thrombocytopenic purpura
2. Vascular: portal vein obstruction, Budd-Chiari syndrome, haemoglobinopathies (Sickle-cell disease, thalassemia)
3. Infective: AIDS, mononucleosis, septicaemia, tuberculosis, brucellosis, malaria, infective endocarditis
4. Traumatic: haematoma, rupture
5. Autoimmune: rheumatoid arthritis, SLE
6. Metabolic: Gaucher's disease, mucopolysaccharidoses, amyloidosis, Tangier disease
7. Inflammatory: sarcoidosis
8. Neoplastic: CML, metastases, myeloproliferative disorders

PATHOLOGIC BASIS OF DISEASE

Robbins *Pathologic Basis Of Disease* is an internationally used reference textbook in pathology. Its chapters under the heading 'General Pathology' can be considered to be a version of the surgical sieve:

3. Acute and chronic inflammation
4. Tissue Repair
5. Hemodynamic Disorders
6. Genetic Disorders
7. Diseases of Immunity
8. Neoplasia
9. Infectious Diseases
10. Environmental and Nutritional Pathology

Bibliography

- Beegle, Clayton C. (Clayton Clarence) - Dulmage, Howard T.: *Bacillus thuringiensis cultures available from the U.S. Department of Agriculture*_ Agricultural Research Service (Southern Region), U.S. Dept. of Agriculture, New Orleans, La, 2002.
- Brown JR: Human actinomycosis. Human Pathology, 2003.
- Brazier JS and Hall V: Actinomyces. In Principles and Practice of Clinical Bacteriology, Emmerson AM, Hawkey PM and Gillespie SH (eds), Wiley, 2007.
- Bolduan, Charles Frederick, Grund, Marie: *Applied bacteriology for nurses*_, Philadelphia and London :W. B. Saunders Company, 2006.
- Buchanan, Robert Earle: *Agricultural and industrial bacteriology*_, London, D. Appleton and Company, 2001.
- Conn Herbert William: *Bacteria in milk and its products*_, London: Rebman 2003.
- Cayley J, Fotherby K, Guillebaud J et al.: Recommendations for clinical practice: Actinomyces like organisms and intrauterine contraceptives. British Journal of Family Planning, 2008.
- Conn, H. W. (Herbert William): *Agricultural bacteriology : a study of the relation of bacteria to agriculture, with special reference to the bacteria in the soil, in water, in the dairy, in miscellaneous farm products, and in plants and domestic animal*, Philadelphia :P. Blakiston's son & co, 2001.
- Conn, H. W. (Herbert William): *Bacteria, yeasts and molds in the home*_, Boston, Ginn and Company, 2003.
- Delepine, Sheridan: *The bacteriological diagnosis of certain infectious diseases in connexion with public health work*, Lancet Office london, 2008.
- Dent VE and Williams RAD: Actinomyces slackii sp. nov. from dental plaque of dairy cattle. International Journal of Systematic Bacteriology, 2006.
- Dobell, Clifford, - Leeuwenhoek, Antoni van: *Antony van Leeuwenhoek and his "Little animals"; being some account of the father of protozoology and bacteriology and his multifarious discoveries in these discipline*, New York, Harcourt, Brace and company, 2002.
- Eastwood, Arthur - Griffith, F - Scott, W. M: *Bacteriological studies*, London H.M. Stationery Off., 2002.
- Foster, N. M. - Sullivan, J. F. - Thompson, C. H.: *Artifacts in the propagation of strains of chronic respiratory disease*

- agents and strains of turkey sinusitis agents in vitro*, National Agricultural Library, Washington, 2004.
- Frankland, Percy: *Bacteria in daily life*, Bombay, Longmans, Green, and co. 2003.
- Funke G, Alvarez N, Pascual C et al.: *Actinomyces europaeus* sp. nov., isolated from human clinical specimens. *International Journal of Systematic Bacteriology*, 2007.
- Gillespie Alexander Lockhart: *The bacteria of the stomach*, Edinburgh, 2002.
- Gupta PK, Erozan YS and Frost JK: Actinomycetes and the IUCD: an update. *Acta Cytologica*, 2008.
- Gupta PK, Hollander DH and Frost JK: Actinomycetes in cervicovaginal smears: an association with IUD usage. *Acta Cytologica*, 2006.
- Hicks, W. B. (William Brooks), Jenkins, M. K. (Minnie Kimball), Pennington, M. E. (Mary Engle), St. John, E. Quintard: *A Bacteriological and chemical study of commercial eggs in the producing districts of the central West*, Washington, D.C. :U.S. Dept. of Agriculture, 2014.
- Hall V and Talbot P: Evaluation of the E-test for susceptibility testing of *Actinomyces* species. *Reviews in Medical Microbiology*, 2007.
- Hoyles L, Falsen E, Foster G et al.: *Actinomyces canis* sp. nov., isolated from dogs. *International Journal of Systematic and Evolutionary Microbiology*, 2000.
- Hall V, Collins MD, Hutson R, Falsen E and Duerden BI: *Actinomyces cardiffensis* sp. nov., from human clinical sources. *Journal of Clinical Microbiology*, 2002.
- Lehmann Karl Bernhard: *Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik*, Munchen :Lehmann, 2006.
- Lehmann, K. B. (Karl Bernhard), Neumann, R. O.: *Atlas and principles of bacteriology and text-book of special bacteriologic diagnosis*, Philadelphia :W.B. Saunders, 2001.
- Lehmann, Karl Bernhard, Neumann, Rudolf Otto, Weaver, George H.: *Atlas and principles of bacteriology and text-book of special bacteriologic diagnosis*, Philadelphia and London, W. B. Saunders & company, 2001.
- Lewis C. J: *The bacteriological examination of the renal secretion in certain of the zymotic diseases, with subsidiary differential experiments*, [S.l.] :[s.n.], 2001.
- Lee YC, Min D, Holcomb K et al.: Computed tomography guided core needle biopsy diagnosis of pelvic actinomycosis. *Gynecological Oncology*, 2000. Traynor RM, Parratt D, Duguid HLD and Duncan ID: Isolation of actinomycetes from cervical specimens. *Journal of Clinical Pathology*, 2001.
- Mace, Eugene: *Atlas de microbiologie*, Paris :J.-B. Baillière, 2008.
- Magnin, Antoine, Sternberg, George Miller: *The Bacteria*, Boston, Little, Brown, and Company, 2000.
- Moor C. G. - Pearmain T. H.: *Applied bacteriology; an introductory handbook for the use of students, medical officers of health, analysts and sanitarians*, London : Bailliere, Tindall and Cox, 2008.
- Newman, George: *Bacteria, especially as they are related to the economy of nature, to industrial processes, and to the public health*, New York, Putnam, 2009.
- Pearmain, T. H.: *Applied bacteriology, an introductory handbook for the use of students*, London, Bailliere, Tindall & Cox, 2007.
- Pearmain, Thomas Hames: *Applied bacteriology; an introductory handbook for the use of students, medical officers of health, analysts and sanitarians*, London, Baillière, Tindall and Cox, 2007.
- Reid, St. George: *Bacteriological diagnosis*, London: Bailliere, Tindall, and Cox, 2007.
- Savage, William G. (William George): *The bacteriological examination of food and water*, Cambridge [Eng.] University Press, 2006.
- Schneider, Albert: *Bacteriological methods in food and drug laboratories, with an introduction to micro-analytical methods*,

- Philadelphia, P. Blakiston's son & co., 2005.
- Slater, Charles, Spitta, Edmund Johnson: *An atlas of bacteriology : containing one hundred and eleven original photomicrographs with explanatory text*, London : Lippincott, 2008.
- Smith, Erwin F. (Erwin Frink:) *Bacteria in relation to plant diseases*, Washington, D.C., Carnegie Institution of Washington, 2005.
- Silver JG, Martin AW and McBride BC: Experimental transient bacteraemias in human subjects with clinically healthy gingivae. *Journal of Clinical Periodontology*, 2009.
- Slack JM and Gerencser MA: *Actinomyces, Filamentous Bacteria. Biology and Pathogenicity*, Burgess, Minneapolis, 2005.
- Smego RA and Foglia G: *Actinomycosis*. *Clinical Infectious Diseases*, 2008.
- Sternberg, George Miller, Magnin, Antoine: *Bacteria*, New York : William Wood & Company, 2005.
- Vandamme P, Falsen E, Vancanneyt M et al.: *Characterization of Actinomyces turicensis and Actinomyces radingae strains from human clinical samples*. *International Journal of Systematic Bacteriology*, 2008.
- Will Corporation: *Bacteriological apparatus: pathological, biochemical*, Rochester, N. Y.: The Will Corporation, 2000.
- Woodhead, German Sims: *Bacteria and their products*, London: Charles Scribner, 2002. Bhagavan BS and Gupta: *Genital actinomycosis and intrauterine contraceptive devices*. *Human Pathology*, 2008.
- Weese WC and Smith IM: *A study of fifty seven cases of actinomycosis over a thirty six-year period*. *Archives of International Medicine*, 2005.
- Williams CE, Lamb GHR and Lewis-Jones HG: *Pelvic actinomycosis: beware the intrauterine contraceptive device*. *British Journal of Radiology*, 2000.

Index

A

Acquisition 56, 211, 235
Adolescent 50, 216
Alcohol 80, 82, 83, 84, 85, 86, 88, 89, 93,
102, 185, 218, 252,
254, 275, 279
Analysis 17, 24, 36, 37, 59, 67, 88, 89, 90, 152,
154, 161, 183, 188,
189, 190, 210, 212,
213, 214, 215, 219,
223, 227, 243, 245,
251, 253, 258, 259,
264, 268, 269, 270,
271, 272, 273, 287
Archaea 2, 3, 8, 9, 10, 15, 45, 47, 49, 53, 54
Authority 264

B

Bacteria 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 21,
24, 25, 27, 28, 30, 36,
42, 43, 45, 46, 47, 48,
49, 50, 51, 52, 53, 54,
55, 56, 57, 58, 59, 60,
61, 62, 63, 64, 65, 66,
67, 68, 69, 70, 71, 72,
73, 75, 92, 95, 97, 99,
100, 103, 104, 105,
112, 113, 114, 118, 120,
124, 126, 129, 131,

132, 133, 139, 142,
144, 155, 158, 161,
162, 164, 166, 171,
176, 181, 193, 194,
196, 197, 206, 210,
211, 229, 256

Brain Injury 246

C

Cancer 25, 26, 31, 33, 34, 35, 36, 40, 41, 42,
43, 54, 55, 60, 62, 72,
73, 76, 78, 79, 92, 99,
101, 102, 105, 108,
110, 112, 119, 120, 121,
123, 128, 132, 137,
141, 142, 160, 165,
166, 176, 179, 180,
184, 186, 188, 198,
200, 216, 223, 229,
233, 237, 242, 243,
244, 249, 250, 252,
263, 269, 271, 280,
283, 284, 285, 286

Cancer Therapy 166

Chemolithotrophy 6

Cognitive 69, 70, 82, 83, 86, 87, 88, 122, 123,
223, 246, 247, 263,
264, 278, 279, 280

Communication 70, 114, 120, 159, 161, 224,

248, 263, 264, 266,
268, 269, 272, 280,
285
Conclusion 186, 187, 241
Culture 44, 48, 49, 56, 158, 180, 202, 216,
224, 256, 260, 269,
279

D

Development 15, 16, 21, 28, 30, 31, 36, 39,
42, 49, 50, 54, 55, 57,
61, 68, 69, 71, 73, 75,
78, 79, 100, 101, 102,
108, 113, 115, 116, 121,
126, 127, 129, 130,
151, 153, 154, 156,
159, 161, 162, 167,
173, 181, 183, 184,
192, 202, 205, 213,
215, 216, 217, 218,
239, 246, 248, 253,
266, 268, 280
Discovery 11, 12, 15, 25, 104, 113, 205, 216,
282
Diversity 2, 8, 16, 31, 46, 48, 50, 51, 53, 55,
59, 60, 61, 65, 66, 73,
117, 129, 130, 132,
134, 138, 167, 172,
177, 178, 212
DNA 9, 15, 20, 21, 22, 23, 24, 25, 26, 27, 28,
29, 30, 31, 32, 33, 34,
35, 36, 37, 38, 39, 40,
41, 43, 44, 46, 48, 50,
71, 101, 103, 110, 113,
127, 139, 142, 143,
154, 170, 179, 180,
186, 229, 231, 262,
273, 283, 288
DNA Era 22
DNA Technology 228

E

Education 78, 197, 216, 264
Emergence 19, 211, 212, 213
Emphasize 82, 106, 187, 249, 263
Energy 1, 2, 3, 4, 5, 6, 7, 8, 12, 55, 58, 61, 108,
154, 187, 281
Environment 5, 15, 16, 27, 28, 29, 30, 33, 40,
42, 52, 56, 57, 58, 62,
66, 67, 68, 69, 82, 85,
101, 114, 115, 139,

151, 156, 158, 161,
163, 164, 166, 178,
184, 185, 209, 210,
211, 228, 278, 281,
284, 286
Epidemiology 75, 78, 183, 184, 185, 187, 188,
189, 191, 204, 208,
210, 241, 242
Evolution 15, 17, 18, 25, 28, 31, 32, 36, 37, 41,
42, 55, 118, 120, 137,
138, 145, 185, 208,
210, 213, 214, 270

F

Fermentation 2, 3, 5, 54, 55, 57, 58, 60, 63
Fitness 32, 35, 36, 37, 41, 235
Freud 86
Fungi 45, 47, 48, 50, 53, 55, 67, 69, 70, 104,
109, 114, 181, 193,
209

G

Gene 15, 16, 19, 20, 21, 23, 25, 26, 27, 28, 29,
30, 31, 32, 34, 35, 36,
38, 39, 40, 41, 42, 43,
55, 56, 72, 102, 109,
122, 128, 133, 137,
140, 141, 142, 144,
146, 148, 149, 152,
155, 163, 166, 175,
181, 187, 191, 192,
193, 206, 229
Generation 3, 6, 8, 17, 18, 19, 21, 26, 27, 30,
42, 43, 49, 101, 116,
126, 127, 129, 150,
170, 171, 172, 177,
180, 192, 211
Geography 55, 215
Gestalt 242
Growth 2, 4, 7, 9, 10, 39, 42, 43, 49, 50, 51,
52, 55, 56, 58, 59, 62,
63, 64, 65, 68, 69, 71,
72, 79, 103, 105, 106,
107, 108, 114, 120,
121, 122, 125, 137,
162, 163, 164, 165,
166, 168, 208, 215,
216, 259, 260, 280,
283, 284

H

Health Management 188
 History 2, 15, 17, 32, 66, 68, 82, 84, 85, 114,
 158, 184, 185, 187,
 197, 204, 206, 215,
 216, 218, 223, 224,
 225, 227, 228, 233,
 236, 240, 242, 243,
 244, 251, 259, 263,
 265, 273, 277, 278
 HIV 26, 40, 63, 73, 76, 77, 78, 79, 81, 84, 102,
 112, 124, 125, 128,
 129, 132, 135, 136,
 137, 145, 159, 165,
 197, 198, 201, 202,
 203, 204, 205, 206,
 207, 208, 211, 212,
 218, 238, 260, 265
 Hormones 53, 55, 92, 101, 121, 122, 255, 271
 Human Immunity 119, 216
 Hybridization 16, 17, 19, 20, 24, 29
 Hygiene 46, 52, 60, 183, 192, 194, 195, 218,
 232, 262, 275

I

Immune System 31, 46, 47, 49, 50, 55, 57,
 67, 71, 73, 75, 92, 95,
 99, 100, 101, 110, 112,
 113, 114, 115, 116, 117,
 118, 119, 120, 121,
 122, 123, 124, 125,
 126, 127, 129, 130,
 133, 134, 135, 137,
 138, 139, 140, 141,
 142, 144, 145, 146,
 148, 152, 153, 154,
 155, 156, 159, 160,
 162, 164, 167, 170,
 171, 172, 173, 175,
 176, 180, 200, 201,
 207, 261, 282, 287,
 288
 Immunology 101, 112, 113, 120, 123, 135, 136,
 139, 151, 171, 179,
 270, 271, 272
 Inanimate 208
 Influence 15, 27, 30, 32, 51, 54, 56, 60, 61,
 72, 101, 117, 130, 148,
 163, 166, 184, 193,

212, 214, 218, 228,
 277

Inhibition 4, 57, 71, 107, 150, 152, 165
 Intensity 106, 107, 202, 228, 251, 258, 276,
 277

Internalization 283

L

Laboratory 24, 25, 43, 62, 91, 92, 130, 149,
 188, 194, 213, 214,
 225, 232, 235, 242,
 259, 269, 270, 271,
 272, 273

M

Management 76, 82, 89, 90, 188, 204, 209,
 212, 224, 244, 258,
 268, 272, 287
 Medical Laboratory 232, 270, 271, 272, 273
 Medicine 25, 42, 75, 76, 80, 113, 118, 123,
 135, 149, 157, 183,
 184, 185, 191, 192,
 193, 205, 208, 215,
 216, 221, 224, 226,
 227, 228, 240, 243,
 245, 248, 262, 264,
 266, 268, 270, 271,
 287
 Methyloleprophy 3
 Microscopy 24, 109, 243, 255, 256, 257, 261,
 270, 271

N

Nature 2, 3, 6, 8, 15, 16, 19, 21, 28, 29, 30, 37,
 46, 85, 92, 101, 102,
 130, 142, 151, 154,
 158, 187, 215, 222,
 240, 247, 250, 251,
 253, 272, 279
 Network 86, 108, 125, 135, 136, 137, 138, 175,
 243

O

Oral Cavity 47, 64, 67, 68, 69
 Organization 76, 89, 138, 192, 197, 209, 263

P

Parental 211, 217
 Pathology 170, 184, 222, 227, 228, 270, 271,
 272, 273, 279, 285,

288
 Pathophysiology 202, 208, 230, 237, 270
 Perceptual 278
 Pharmacotherapy 82
 Phototrophy 8
 Physiology 16, 25, 46, 71, 106, 107, 135, 157,
 222, 230, 235, 270
 Pregnancy 59, 65, 66, 67, 118, 135, 179, 180,
 200, 201, 202, 204,
 206, 207, 208
 Prenatal Therapy 180
 Psychology 222

R

Radio 228

S

Science 46, 48, 77, 89, 100, 113, 184, 187, 188,
 191, 193, 234, 256,
 270
 Structure 15, 20, 22, 24, 27, 29, 30, 34, 35,
 40, 41, 64, 83, 106,
 113, 119, 139, 142,
 143, 145, 149, 152,
 153, 154, 160, 166,
 170, 172, 174, 181,
 182, 212, 213, 214,
 219, 222, 227, 255,
 269, 281, 282, 287

T

Technique 22, 24, 25, 43, 152, 153, 185, 227,
 229, 245, 246, 247,
 248, 249, 256, 257,
 259, 260
 Technology 36, 39, 43, 141, 225, 228, 234,
 235, 243, 258, 270
 Terminology 19, 45, 76, 140, 273
 Theory 16, 18, 28, 32, 34, 36, 37, 40, 86, 89,
 108, 113, 123, 125,
 134, 135, 136, 137,
 175, 183, 187, 193,
 208, 213, 215, 216,

219, 220, 227, 230
 Treatment 5, 8, 25, 50, 51, 59, 67, 76, 78, 81,
 82, 83, 84, 85, 86, 87,
 88, 89, 90, 103, 107,
 111, 112, 137, 145,
 146, 149, 150, 158,
 166, 179, 180, 183,
 185, 202, 203, 204,
 205, 208, 216, 221,
 222, 224, 225, 226,
 227, 228, 229, 233,
 234, 238, 239, 240,
 241, 243, 244, 245,
 247, 248, 249, 250,
 251, 252, 254, 257,
 259, 260, 262, 263,
 264, 269, 270, 274,
 279, 287, 288

Tumor 26, 35, 43, 62, 73, 92, 94, 95, 101,
 103, 104, 108, 110,
 115, 116, 120, 121,
 122, 127, 128, 132,
 141, 153, 155, 158,
 162, 165, 166, 233,
 243, 250, 280, 283,
 284, 285, 286

Tumor Immunology 120

U

Ulcers 60, 197, 237

V

Virus 30, 31, 37, 47, 63, 73, 102, 103, 113, 116,
 120, 124, 128, 129,
 135, 137, 141, 145,
 155, 159, 165, 169,
 176, 179, 181, 194,
 195, 198, 200, 201,
 204, 205, 206, 208,
 209, 211, 212, 215,
 229, 238, 259, 260,
 261, 265