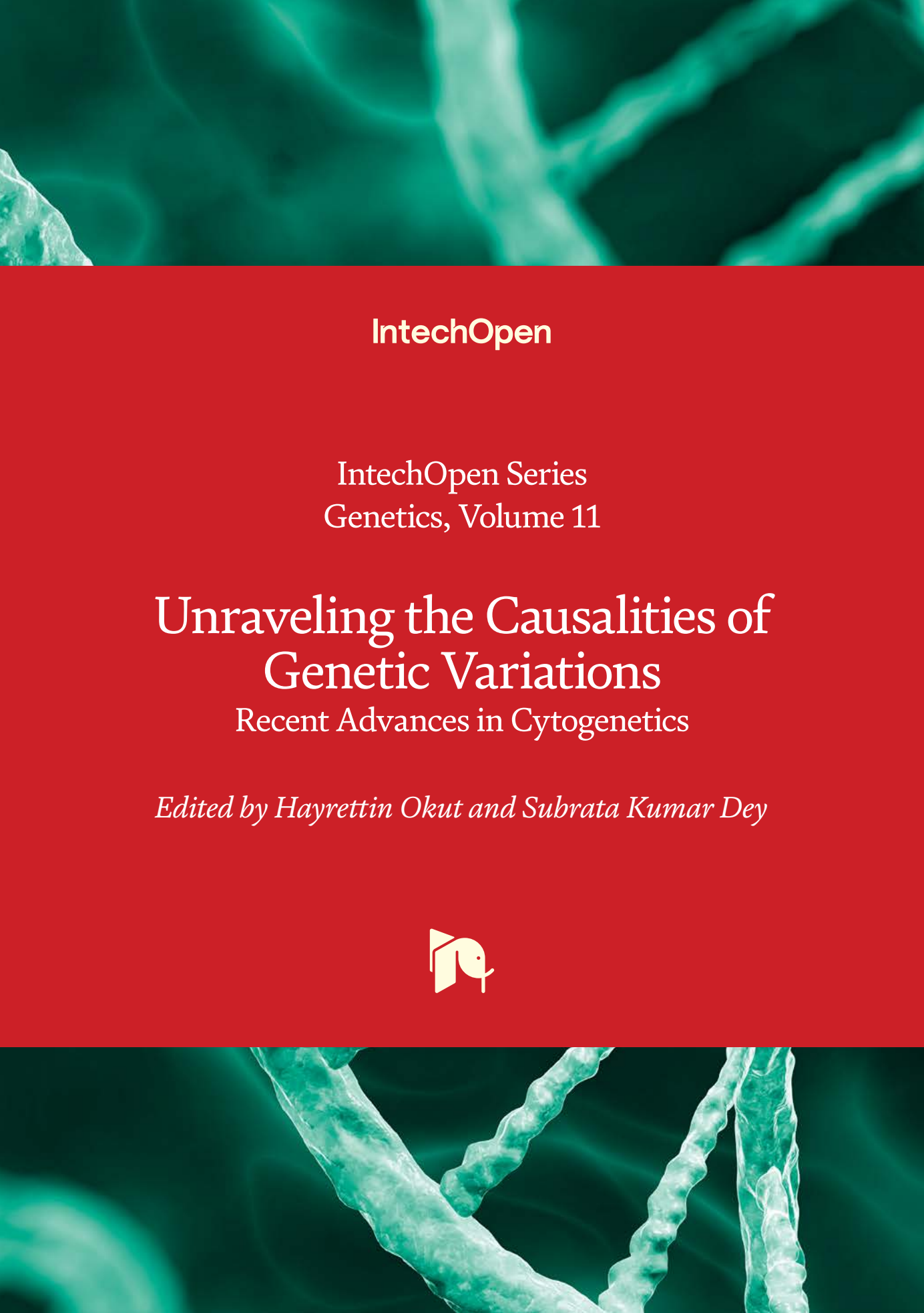




IntechOpen

IntechOpen Series
Genetics, Volume 11

Unraveling the Causalities of Genetic Variations

Recent Advances in Cytogenetics

Edited by Hayrettin Okut and Subrata Kumar Dey



Unraveling the Causalities of Genetic Variations - Recent Advances in Cytogenetics

*Edited by Hayrettin Okut
and Subrata Kumar Dey*

Published in London, United Kingdom

Unraveling the Causalities of Genetic Variations - Recent Advances in Cytogenetics

<http://dx.doi.org/10.5772/intechopen.111119>

Edited by Hayrettin Okut and Subrata Kumar Dey

Contributors

Aditi Chaudhary, Aram Asareh Zadegan Dezfuli, Asra Kirmani, Chanda Siddoo-Atwal, Ching-Fen Hsu, Faisal Ahmed, Hayrettin Okut, Jaffar Abass Wani, Junyi Xin, Kachienga Leonard, Magwalivha Mpho, Mahrukh Hameed Zargar, Manikantan Pappuswamy, Meilin Wang, Mohammad Hashemzadeh, Mulong Du, Nethathe Bono, Niranjana Karukayil Gopalakrishnan, Qian Jiang, Rikhotso Mpumelelo Casper, Rosey Zackula, Sana Syed Ansar, Shi-Yu Rao, Silu Chen, Subrata Kumar Dey, Tahir Mohiuddin Malla, Xiaoling Guo, Xin Zheng, Xiu-Zhi Fu, Yuqing Ran, Zafar A. Shah

© The Editor(s) and the Author(s) 2025

The rights of the editor(s) and the author(s) have been asserted in accordance with the Copyright, Designs and Patents Act 1988. All rights to the book as a whole are reserved by INTECHOPEN LIMITED. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECHOPEN LIMITED's written permission. Enquiries concerning the use of the book should be directed to INTECHOPEN LIMITED rights and permissions department (permissions@intechopen.com).

Violations are liable to prosecution under the governing Copyright Law.



Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 4.0 License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be found at <http://www.intechopen.com/copyright-policy.html>.

Notice

Statements and opinions expressed in the chapters are those of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in London, United Kingdom, 2025 by IntechOpen

IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 167-169 Great Portland Street, London, W1W 5PF, United Kingdom

For EU product safety concerns: IN TECH d.o.o., Prolaz Marije Krucifikse Kozulić 3, 51000 Rijeka, Croatia, info@intechopen.com or visit our website at intechopen.com.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Unraveling the Causalities of Genetic Variations - Recent Advances in Cytogenetics

Edited by Hayrettin Okut and Subrata Kumar Dey

p. cm.

This title is part of the Genetics Book Series, Volume 11

Topic: Genomics

Series Editor: Kenji Ikehara

Topic Editor: Irina Vlasova-St. Louis

Print ISBN 978-1-83769-903-2

Online ISBN 978-1-83769-902-5

eBook (PDF) ISBN 978-1-83769-904-9

ISSN 3049-7094

If disposing of this product, please recycle the paper responsibly.

IntechOpen Book Series

Genetics

Volume 11

Aims and Scope of the Series

“Genetics,” which has been proud of its tradition since Mendel presented his research results in 1865, initially progressed quite slowly due to simple observational approaches of individuals and groups. However, the discovery of double-stranded DNA by Watson and Crick about 70 years ago triggered rapid progress in life sciences, including genetics, which was primarily conducted using *Escherichia coli* and bacteriophages infecting *E. coli*. Subsequently, genetics has achieved remarkable developments, such as understanding genetic disorders, including cancers, through research on the biogenesis and differentiation of plants and animals. The two topics of this book series - Human Genetics, and Genomics - will address important areas of advancement in genetics.

Human Genetics: After fundamental genetics, initially studied with the main goal of revealing the functions of individual genes and proteins, genetics expanded from understanding the genetic system itself to understanding many infectious diseases caused by bacteria and viruses. Consequently, human beings are now overcoming infectious diseases by developing medicinal chemicals, including antibiotics and vaccines. However, genetic disorders remain challenging to cure up to now. Nevertheless, even the cure for them, including various cancers, is coming closer to reality due to the rapid progress of human genetics. In this way, the welfare of human life continues to improve, and even longevity, which was once a dream, has been achieved to some extent in recent years.

Genomics: On the other hand, the understanding of the comprehensive interrelationship of whole genes or whole proteins functioning in one organism has become possible now, as research has entered the era of genomics, owing to the rapid progress of base sequence analysis and bioinformatics. The development of genomics has further made it possible to understand the evolutionary processes of organisms through comparative studies among the genomes of many organisms.

This book series will discuss the findings obtained during the advancement of human genetics and genomics. It is also expected that this series will trigger the formation of a better world composed of human beings and all other organisms on Earth through discussions of research results obtained under the development of general genetics.

Meet the Series Editor



Kenji Ikehara graduated from the Department of Industrial Chemistry, Faculty of Engineering, Kyoto University in 1968. He received his B. Eng. (1968) and subsequently earned M. Eng. (1970) and D. Eng. (1976) degrees from Kyoto University. He began his career as a research associate in the Faculty of Science at the University of Tokyo before moving on to become an associate professor in the Faculty of Science at Nara Women's University. He was later promoted to professor and subsequently served as the dean of the Faculty of Science at Nara Women's University. Additionally, he held the position of director at the Nara Study Center of the Open University of Japan. For approximately 15 years, he focused his research on sporulation initiation of *Bacillus subtilis*. Later, he shifted his focus to the origins and evolutionary processes of microbial genes, the genetic code, proteins, and life. He has proposed several hypotheses, including the GC-NSF(a) hypothesis on the origin of genes, the GNC-SNS hypothesis on the genetic code, the protein 0th-order structure hypothesis on the origin of proteins, and the [GADV]-protein world hypothesis (GADV hypothesis) on the origin of life. Furthermore, he served as the local chair of the International Conference, Origin 2014, held in Nara in 2014.

Meet the Volume Editor



Hayrettin Okut, Ph.D., is a Research Professor at the University of Kansas School of Medicine–Wichita. He earned his doctorate in Biostatistics/Biometry in 1992 and previously served as a Senior Research Associate at Wake Forest University’s Center for Genomics and Personalized Medicine. His career includes extensive teaching and collaborations with institutions such as the University of Wisconsin–Madison, Oregon Research Institute, and the University of Oregon. Dr. Okut completed postdoctoral research at the University of Nebraska as a TÜBİTAK scholar and at the University of Oregon as a Fulbright Scholar. Since 2011, he has pioneered the application of Artificial Neural Networks in genomic prediction, publishing numerous peer-reviewed papers and book chapters, and presenting widely at international conferences.



Prof. Dr. Subrata Kumar Dey joined the faculty of Maulana Abul Kalam Azad University of Technology (formerly known as West Bengal University of Technology) as Professor of Biotechnology in 2005. His laboratory has long been involved in research on the molecular genetics of Down syndrome, Alzheimer’s disease, and congenital heart disease. He has done extensive research on radiation-induced genomic instability, radioprotection, and stem cell biology. Prof. Dey received the Golden Jubilee Award for excellence in teaching and research. He has published over 100 research papers in peer-reviewed journals, edited six books on Down syndrome and other congenital disorders, and completed eleven research projects funded by the Government of India. Several students also received a Ph.D. under his supervision. Prof. Dey established a Centre for Genetic Studies to provide diagnostic and counselling services to individuals affected by genetic disorders. Along with teaching and research, Prof. Dey has successfully handled several administrative assignments, taking a dedicated and innovative approach with high integrity. Major administrative assignments handled include Director of the School of Biotechnology and Biological Sciences, Founder Co-ordinator of the Centre for Genetic Studies, Member-Coordinator of the Ekta Incubation Centre, Pro-Vice Chancellor and Vice-Chancellor of Maulana Abul Kalam Azad University of Technology, and Vice Chancellor of Brainware University. Currently, Prof. Dey serves as Vice Chancellor of Swami Vivekananda University, West Bengal, India.

Contents

Preface	XIII
Section 1	
Foundations and Conceptual Advances	1
Chapter 1	3
Introductory Chapter: The Evolving Landscape of Genetic Variation Research <i>by Hayrettin Okut</i>	
Chapter 2	9
Cytogenetics and Microdeletion Syndromes <i>by Mahrukh Hameed Zargar, Tahir Mohiuddin Malla, Faisal Ahmed, Zafar A. Shah, Jaffar Abass Wani and Asra Kirmani</i>	
Chapter 3	19
Recent Molecular Techniques in Cytogenetics <i>by Rikhotso Mpumelelo Casper, Kachienga Leonard, Magwalivha Mpho and Nethathe Bono</i>	
Chapter 4	33
Technological Advances in Cancer Cytogenetics <i>by Subrata Kumar Dey</i>	
Section 2	
Methods for Understanding Genetic Causality	43
Chapter 5	45
Discovering Causality with Mendelian Randomization: What It Is and How to Do It <i>by Hayrettin Okut and Rosey Zackula</i>	
Chapter 6	67
Environmental Pollutant-Based Polygenic Risk Score: A Genetic Surrogate for Environmental Pollutants Assessment <i>by Mulong Du, Junyi Xin, Silu Chen and Meilin Wang</i>	

Chapter 7	85
Genetic Changes, Epigenetic Modifications and Antibiotic Resistance Caused by Bacterial Infection	
<i>by Mohammad Hashemzadeh and Aram Asareh Zadegan Dezfuli</i>	
Section 3	117
Applications and Case Studies in Human Disease	
Chapter 8	119
Asymmetry of Brain and Behavior of People with Williams Syndrome in Verbal and Nonverbal Processing	
<i>by Ching-Fen Hsu, Qian Jiang, Shi-Yu Rao, Xiu-Zhi Fu and Xin Zheng</i>	
Chapter 9	129
iPS Cell Models of Human Single Gene Inheritance Diseases	
<i>by Yuqing Ran and Xiaoling Guo</i>	
Chapter 10	141
Comparative Analysis of Carcinogenic Compounds in Cooking Oil Emissions and Cigarette Smoke: Implications for Genetic Disorders and Cancer Susceptibility	
<i>by Niranjana Karukayil Gopalakrishnan, Sana Syed Ansar, Manikantan Pappuswamy and Aditi Chaudhary</i>	
Chapter 11	155
β -Thalassemia in India: A Classic Case of Genetic Adaptation in Response to a Pathogen with a Historical Perspective on Ancient Population Migrations (a Mini-Review)	
<i>by Chanda Siddoo-Atwal</i>	

Preface

The rapid advancement in biomedical research in human genetics has transformed our understanding of genetic variation, disease risk, and treatment response. This edited volume, “*Unraveling the Causalities of Genetic Variations – Recent Advances in Cytogenetics*”, presents a comprehensive overview of genetic variation research by current experts in the field.

Genetic differences, ranging from single-nucleotide changes to large chromosomal rearrangements, form the basis of human diversity and are key to understanding the mechanisms of complex diseases. Cytogenetics has undergone a dramatic shift with advances in technology, enabling more in-depth explorations of how genetic differences impact health outcomes. This book reflects the status of this evolving field and offers glimpses into future uses.

The chapters explore genetic variation research in a step-by-step fashion, beginning with its foundation and progressing to applications in personalized medicine. The chapters describe the evolving landscape of genetic studies, indicating a shift from single-gene diseases to research that encompasses multiple genetic elements, environmental factors, and epigenetic modifications. Population genetics observations describe variations that contribute to adaptation and evolution, such as the protective effect of β -thalassemia genes against malaria in specific populations.

The book emphasizes methodological advances through coverage of environmental pollutant-based polygenic risk scores (EpPRS) and polygenic risk scores (PRS). These methods combine genetic and environmental data to improve predictive risk assessment. Microdeletion syndromes are also elaborated upon. Chapters outline how array comparative genomic hybridization and next-generation sequencing have revolutionized the identification and characterization of these disorders.

Translation of genetic variation research into personalized medicine is the primary emphasis within the field. Chapters demonstrate how genetic profiling can be utilized to inform personalized prevention and treatment, which has the potential to revolutionize the practice of medicine. The book addresses challenges facing the contemporary world, for example, the genetic causes of antibiotic resistance. It demonstrates how bacteria evolve their genetic material to develop resistance.

The book concludes with talks on technological advancements that will characterize the field of cytogenetics. Long-read sequencing (sequencing large portions of DNA at once), optical mapping (large-scale imaging of DNA structure), and single-cell analysis (analysis of genetic variation in single cells) now enable the exploration of complex variation with unerring accuracy. Methods such as Mendelian randomization (pointing towards causality via genetic variants) can be employed to determine causal relationships between variations and health status. Throughout, care is taken of the ethical, legal, and social dimensions of genetic research.

This book integrates existing knowledge and highlights emerging research areas. The varying expertise of the authors provides comprehensive coverage of this developing branch. The authors' vision provides readers with both the fundamental knowledge and new developments to construct the future of genetic medicine.

My sincere gratitude goes to my colleague, Rosey Zackula (University of Kansas, School of Medicine, Wichita), who has always stood by me for many things, including exchanging scholarly ideas. I would also like to thank Antonija Grgec, a Publishing Process Manager at IntechOpen, who has always been supportive of overcoming formalities and keeping me informed through her instructions and alerts. Lastly, we thank all the contributors who have made this comprehensive evaluation of genetic variation studies achievable through sharing their experience and knowledge.

Hayrettin Okut, Ph.D.

Department of Population Health and Office of Research,
University of Kansas,
School of Medicine-Wichita,
Wichita, KS, USA

Subrata Kumar Dey, Ph.D.

Swami Vivekananda University,
Barrackpore, West Bengal, India

Section 1

Foundations and Conceptual Advances

Chapter 1

Introductory Chapter: The Evolving Landscape of Genetic Variation Research

Hayrettin Okut

1. Introduction

The field of biomedical research in human genetics is advancing rapidly. Because genetic variations affect biological functions, disease susceptibility, and responses to environmental factors, it is crucial to investigate these variations, which are defined as changes in DNA sequences. Human diversity is supported by genetic variations, which range from single-nucleotide changes to significant chromosomal alterations. In some instances, these variations indicate the risk of disease. Cytogenetics has undergone significant changes as a result of recent technological advancements, supporting a more comprehensive examination of the health effects associated with genetic variation.

Genetic research has shifted from focusing solely on single-gene disorders to examining health risks associated with multiple genes and environmental factors. We now understand that genetic variations operate within complex networks influenced by the environment, epigenetic changes, and random events. As Lappalainen et al. [1] note, epigenetic changes make it even more challenging to understand how genes and the environment interact to affect health. This broader approach lets researchers study how genetic variations contribute to disease.

Advanced technologies, including genome-wide studies, next-generation sequencing, and sophisticated computational tools, are utilized in modern cytogenetics. By identifying genetic variations across large populations, high-throughput sequencing advances our understanding of disease risk and genetic diversity. New genetic markers, improved risk prediction, and enhanced individualized treatment are all made possible by advanced data analysis techniques that facilitate the identification of significant genetic alterations and their associated consequences.

2. Discovering genetic variations

Genetic variations come in many forms and can have different effects. Some have little impact, while others can strongly affect how the body works. To better understand these effects, researchers combine genetic data with information about proteins, metabolism, and gene activity. This combined approach helps show how genetic changes influence health.

Studying genetic variations helps us learn more about how humans adapt and evolve. For example, some Indian populations are protected against malaria because

they have the β -thalassemia trait, showing how migration and natural selection influence our genetic makeup. Analysis of these adaptations enhances understanding of human evolutionary processes and informs strategies for addressing future health challenges.

Population genetics examines how genetic variation is distributed among different groups of people. By examining specific genetic variants, scientists can determine how people migrated in the past, identify signs of natural selection, and describe adaptations unique to certain populations. These findings help create tailored preventative initiatives and make it easier to understand how diseases spread. Understanding people's genetic histories may help healthcare to be more equitable and effective.

Polygenic risk scores (PRSs) are a novel approach to assessing an individual's likelihood of developing a disease and understanding how environmental factors may impact their health. For example, Environmental Pollutant-based Polygenic Risk Scores (EpPRS) combine genetic and environmental information to estimate a person's exposure to certain pollutants. PRS methods examine multiple genetic variants to provide a comprehensive view of disease risk. Although there are significant challenges in utilizing PRS with diverse populations, there are also challenges in connecting these scores to widespread health databases.

Microdeletion syndromes are a group of genetic disorders that are notable because they can present in several ways. They are caused by the loss of specific parts of chromosomes. These disorders frequently display complex symptoms that affect multiple organ systems, as they have lost various genes in the deleted area. Modern cytogenetic methods, including array comparative genomic hybridization and next-generation sequencing, have revolutionized the detection and characterization of microdeletions. These methods facilitate the connection between genotype and phenotype, enabling more effective patient management.

3. Closing the gap between personalized medicine and genetic adaptations

Personalized medicine is one of the most groundbreaking applications of research into genetic variation. By tailoring prevention and treatment approaches to a person's genetic profile, it promises to change how healthcare is delivered and improve patient outcomes. Jamalnia and Weiskirchen [2] suggest that combining genetic data with lifestyle and environmental data enables more targeted interventions, which in turn improve treatment efficacy and reduce adverse effects. To make personalized medicine effective, we need to ensure that everyone has access to genetic testing, develop clinical decision support tools, and educate healthcare professionals about the principles of genetic medicine. Induced pluripotent stem (iPS) cell models for single-gene inheritance illnesses are one way that personalized medicine can change how we think about genetic disorders. These models are great for studying how diseases work and testing personalized treatments.

Research on antibiotic resistance highlights the importance of understanding how bacteria modify their genetic material. Antibiotic-resistant bacteria pose a significant health threat because they often develop resistance through mutations and genetic exchange [3]. Researchers can develop more effective plans to prevent infections and maintain public health by gaining a deeper understanding of these processes.

Understanding bacterial genetic alterations is crucial, as evidenced by research on antibiotic resistance. Because they frequently develop resistance through mutations and genetic exchange, antibiotic-resistant strains represent a serious health risk.

By understanding these processes, researchers can develop more effective plans to combat infections and protect public health.

Research on the carcinogens found in cigarette smoke and cooking oil fumes shows how the environment and genetics interact to influence cancer risk. Genetic variations can affect how toxins are processed or how DNA is repaired, which can impact a person's vulnerability to cancer brought on by these exposures. This highlights the importance of considering both environmental and genetic factors when evaluating cancer risk.

4. The cutting edge of cytogenetics: Advancements in technology and prospects

New technologies that help us discover and explain genetic variants are advancing cytogenetics rapidly. Advanced sequencing, specialized staining, and detailed chromosome analysis enable us to study genetic variations more closely. Using these methods with computer tools also makes predictions and diagnoses more accurate.

Long-read sequencing and optical mapping enable the study of repetitive DNA and complex genetic changes. Single-cell sequencing shows changes in specific cell types and highlights variations between individual cells. As these techniques continue to improve, we will learn even more about how genetic variations affect health.

Combining cytogenetics with fields like functional genomics and systems biology is opening new paths for genetic research. Using different types of data helps us understand how genetic changes work. Advanced computer methods, including artificial intelligence and machine learning, help find important patterns and possible treatment targets.

For example, Mendelian randomization (MR) is an epidemiological technique that aims to overcome the limitations of conventional epidemiological studies and the practical difficulties associated with randomized controlled trials (RCTs) by leveraging genetic variation to detect causal effects between an exposure and an outcome. MR utilizes genetic differences as instrumental variables to overcome the limitations of observational research and strengthen the case for causality. It is expected that MR approaches will become more useful in clinical decision-making and biomedical research as they continue to improve, especially for complex features [4].

As genetic research advances, we must consider its ethical, legal, and social implications. Issues such as consent, privacy, discrimination, and fair access to testing require ongoing attention. By addressing these concerns, we can ensure that genetic research benefits people and respects their rights.

5. Conclusion: Toward a precision health future

Interdisciplinary collaboration across scientific fields has yielded novel insights into health and disease. Integrating genetic, environmental, and clinical data enhances precision health, which includes public health, disease prevention, and personalized medicine. This integration enables earlier identification of health risks and supports the selection of more effective treatment strategies.

Analysis of genetic variation improves the diagnosis and treatment of diseases. Furthermore, incorporating clinical information with genetic test results enables more precise diagnoses and accelerates earlier initiation of treatment. Additionally,

understanding the mechanisms of genetic diseases informs drug development and supports the selection of optimal therapies for individual patients. While gene editing holds potential for correcting deleterious genetic mutations, concerns regarding safety and equitable access persist.

Recognizing the complexity of genetic variations and their causal relationships to human health is crucial for the advancement of cytogenetics. Maximizing the impact of genetic variation research depends on interdisciplinary approaches that integrate clinical medicine, environmental science, and genetics. Effective translation of genetic discoveries into improved health outcomes requires sustained collaboration among scientific disciplines, as noted by Heng et al. [5]. In a globally interconnected context, this comprehensive strategy enables the effective management of complex health challenges and facilitates the application of genetic variations to improve medical interventions.

The study of genetic variations is key to modern biomedical research. It helps us understand how diseases start, how our bodies work, and how to treat illnesses. As genetic techniques improve, we are moving closer to utilizing genetic data for more effective treatments, prevention, and informed health choices. To achieve this goal, we must continue to keep innovating, working together, and translating discoveries into real health benefits.

Author details

Hayrettin Okut

Department of Population Health and Office of Research, University of Kansas
School of Medicine, Wichita, KS, United States

*Address all correspondence to: hokut@kumc.edu

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Lappalainen T, Li YI, Ramachandran S, Gusev A. Genetic and molecular architecture of complex traits. *Cell*. 2024;**187**(5):1059-1075. DOI: 10.1016/j.cell.2024.01.023
- [2] Jamalnia M, Weiskirchen R. Advances in personalized medicine: Translating genomic insights into targeted therapies for cancer treatment. *Annals of Translational Medicine*. 2025;**13**(2):18. DOI: 10.21037/atm-25-34. Epub 2025 Apr 29
- [3] Oliveira M, Antunes W, Mota S, Madureira-Carvalho Á, Dinis-Oliveira RJ, Dias da Silva D. An overview of the recent advances in antimicrobial resistance. *Microorganisms*. 2024;**12**(9):1920. DOI: 10.3390/microorganisms12091920
- [4] Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. *Nature Reviews Methods Primers*. 2022;**2**:6. DOI: 10.1038/s43586-021-00092-5
- [5] Heng E, Thanedar S, Heng HH. Challenges and opportunities for clinical cytogenetics in the 21st century. *Genes*. 2023;**14**(2):493. DOI: 10.3390/genes14020493

Chapter 2

Cytogenetics and Microdeletion Syndromes

*Mahrugh Hameed Zargar, Tahir Mohiuddin Malla,
Faisal Ahmed, Zafar A. Shah, Jaffar Abass Wani
and Asra Kirmani*

Abstract

The emergence of molecular cytogenetic technologies has transformed the process of identifying new microdeletion syndromes. Traditional methods of identifying these syndromes relied on the chance discovery of patients exhibiting specific clinical features and visible chromosomal rearrangements through techniques like G-banding. However, with the introduction of comparative genomic hybridization using microarrays, it has become possible to detect previously unknown and recurring imbalances in individuals presenting with conditions such as mental retardation and nonspecific features. In contrast to the conventional “phenotype-first” approach in cytogenetics, array-based comparative genomic hybridization allows for a “genotype-first” approach. This means that the focus is on identifying genetic variations or imbalances in the genome before associating them with specific clinical phenotypes. By facilitating the identification of novel genomic disorders and recurrent imbalances, this technology has revolutionized the field of cytogenetics and deepened our understanding of the genetic underpinning of microdeletion syndromes. The present chapter emphasizes the role of cytogenetics technologies in delineating the aetiology of microdeletion syndromes.

Keywords: cytogenetics, microdeletion syndromes, fluorescent *In situ* hybridization, comparative genomic hybridization, chromosomal microarray

1. Introduction

Cytogenetic analysis plays an important role in delineating the aetiology of numerous suspected chromosomal disorders. Cytogenetics examines the connection between chromosomal abnormalities and genetic disorders in humans [1]. Microdeletion syndromes are a group of disorders caused by the deletion of small chromosomal segments (< 5 Mb in size) that contain multiple genes, each potentially contributing independently to the disease phenotype. These disorders usually arise from the haploinsufficiency of critical genes in the affected region [2].

Microdeletion syndromes often occur when a chromosomal segment is lost, resulting in a deficiency of a small number of crucial genes or, in some instances, a

single gene [3]. Microdeletions are mostly spontaneous and manifest in about 5% of individuals with unexplained cognitive impairment [4, 5]. They are often linked to various congenital abnormalities and developmental delays [3]. The prevalent microdeletion disorders are DiGeorge syndrome (22q11.2), Prader-Willi syndrome, Angelman syndrome (15q11–13), Williams syndrome (7q11.23), and Wolf-Hirschhorn syndrome (4p16.3) [6].

The genetic changes associated with microdeletion syndromes are often undetectable by conventional or high-resolution karyotyping (2–5 Mb), thereby necessitating the use of molecular cytogenetic techniques such as fluorescence *in situ* hybridization (FISH) or the latest array comparative genomic hybridization (CGH) technique [7]. Over the years, cytogenetic analysis has significantly advanced and become a routine part of laboratory testing, providing crucial diagnostic and prognostic information for human diseases. The integration of microarray technology with molecular cytogenetics and traditional chromosome analysis has revolutionized clinical cytogenetics. The benefits of microarray technologies are now widely recognized in the medical and laboratory communities involved in genetic diagnosis, leading to greatly enhanced visualization and validation capabilities [8].

2. Conventional cytogenetics

Banded karyotyping, using conventional methods, is widely accepted as the most reliable and accurate technique for diagnosing and predicting the outcome of genetic disorders. It has been used to scan for changes in the genome that include both increases and decreases in specific parts of the genome, as well as rearrangements within and between the chromosomes [9]. Karyotyping research has been used to establish the causal relationship between certain chromosomal abnormalities and clinical disorders such as congenital malformations, developmental delay (DD), and mental retardation (MR) [10]. Nevertheless, banding analysis is regarded as a process that requires a significant amount of time and manual effort. Typically, it takes around two weeks to get the findings, although there is a possibility that some clinically significant chromosomal abnormalities at a submicroscopic level may go unnoticed [11]. Moreover, the identification of microdeletions and duplications that lead to notable clinical problems including congenital abnormalities, mental retardation (MR), developmental delay (DD), or intellectual impairment (ID) is challenging [12]. The lack of precise bands resolution on particular chromosomes hinders the accurate characterization and interpretation of intricate and hidden chromosomal changes.

3. Molecular cytogenetics

Fluorescence *in situ* hybridization (FISH) and comparative genomic hybridization (CGH) have developed as effective diagnostic methods to address the limitations of banding analysis in molecular cytogenetics. These approaches are often used as supplementary methods to standard cytogenetics for the detection of chromosomal abnormalities [9]. It is often used in addition to traditional approaches to detect chromosomal abnormalities [13–17]. Variable molecular cytogenetic techniques are considered valuable additions or potential alternatives to chromosomal banding.

These techniques improve the comprehensive interpretation of numerical and complex chromosome abnormalities by bridging the gap between conventional banding analysis and molecular genetic studies [18–20]. **Table 1** provides a comparison between various cytogenetic techniques.

3.1 Fluorescence in situ hybridization

FISH uses DNA probes that are fluorescently labeled and specific to chromosomal regions. The probes used in FISH testing are replicated segments of genomic DNA. They have the ability to identify certain DNA sequences and generate a fluorescent signal when they bind to their complimentary sequences. This signal stands out against the backdrop of dyed chromosomes, making FISH testing very effective [21] (**Figure 1**). The FISH technique enables the identification of minor genomic changes ranging from 50 Kb to 100 Kb and also allows for the direct observation of these changes in cells that have not been grown [22]. Fluorescence *in situ* hybridization (FISH) is a commonly used method for identifying microdeletions. Fluorescence *in situ* hybridization (FISH) has emerged as the prevailing technique for detecting prevalent microdeletion disorders such as Prader-Willi syndrome, Angelman syndrome, Velocardiofacial (DiGeorge) syndrome, and Williams syndrome. In cases where there is a strong suspicion based on clinical symptoms, the AI system may also identify uncommon syndromes like Wolf-Hirschhorn syndrome and Smith-Magenis syndrome [6].

Techniques	Method	Characteristics	Application	Advantages
Conventional cytogenetics (G-banding)	Lymphocyte culture, Metaphase arrest, G-banding, Microscopy	Special dye generates specific banding pattern for each chromosome	Detection of numerical and structural chromosomal anomalies	Screening of chromosome level abnormalities genome-wide
Fluorescent <i>In Situ</i> Hybridization	Cultured/ interphase lymphocytes, Fluorescent DNA probes, Microscopy	Labelled DNA is used as a probe to search for target sequences in chromosomes	Detects all types of balanced and unbalanced chromosome anomalies	Interphase cytogenetics Possible, Less time consuming
Comparative Genomic Hybridization	Molecular technique	Comparative hybridization of differentially labelled test DNA and reference DNA	Identify biomarkers Gene discovery, functional analysis	Genome-wide screening of genomic anomalies, No need for cell culture
Array Comparative Genomic Hybridization	Molecular technique	Identification of DNA sequences by specific DNA binding proteins in cells	Identification of cryptic rearrangements (aneuploidy, deletions, duplications or amplifications)	High-resolution target-specific detection of gene amplification, submicroscopic information on imbalances

Table 1.
Comparison of conventional cytogenetics and molecular cytogenetic techniques.

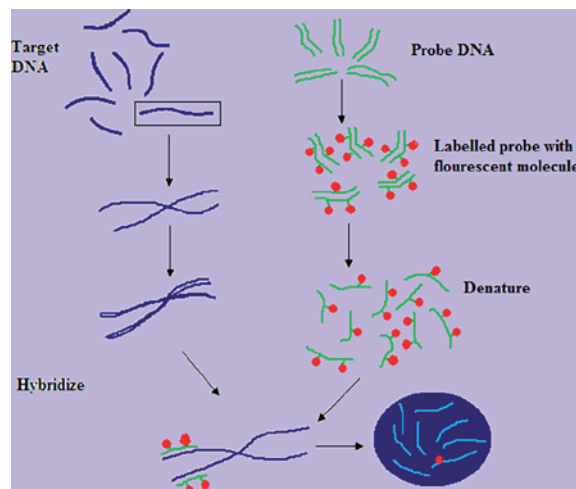


Figure 1.
Schematic diagram of FISH.

3.2 Comparative genomic hybridization

Comparative genomic hybridization (CGH) is a valuable technique for identifying subtle genetic imbalances, such as increases or losses of chromosomal material that may not be easily apparent using standard cytogenetic methods. Comparative genomic hybridization (CGH) enables the identification of amplifications and deletions in specific areas of DNA across all chromosomes. The approach operates by doing a comparison between the genomic content, or DNA, of a patient, or target, and that of a normal control person. The resolution of “classical” or metaphase comparative genomic hybridization (CGH) is quite poor, measuring roughly 15 megabases (Mb) of DNA [10]. Comparative genomic hybridization (CGH) is a technique that may be used to screen the complete genome for abnormalities. It is a modified version of fluorescence *in situ* hybridization (FISH) technology and has the benefit of detecting imbalances over the entire genome [23]. In general, the ability to identify copy number variations using FISH and CGH approaches is only marginally better than traditional karyotyping. Moreover, these tests require a significant amount of manual labour and consume a considerable amount of time, particularly when many genomic areas are examined [24]. In order to identify such irregularities, a method with great resolution is necessary.

3.3 Array comparative genomic hybridization

Array CGH is a kind of CGH where the comparator DNA, RNA, or tissue is arranged in an organized pattern on a glass slide or glass beads. There are three fundamental categories of arrays: bacterial artificial chromosomes (BAC) arrays, oligonucleotide arrays (usually consisting of 60 base pairs), or single nucleotide polymorphism (SNP) arrays (often containing a few nucleotides). Nowadays, most SNP-based arrays include single locus probes in addition to SNPs. These arrays may be either targeted arrays or full genomic arrays. Targeted arrays specifically focus on identifying known microdeletion/microduplication syndromes and other known sites associated with inherited Mendelian illnesses [25]. The first targeted arrays consisted of around 500 to 600 BACs. Whole genomic arrays provide comprehensive coverage

of the complete genome with varied degrees of precision. The first complete genomic arrays consisted of BAC arrays, including around 2600 BACs, evenly distributed at intervals of approximately 1 megabase (Mb) over the whole genome. This arrangement allowed for a resolution of about 1 Mb [26]. Oligonucleotide and SNP arrays have replaced the usage of BACs. The majority of labs use either oligonucleotide or SNP arrays, which typically have an average resolution of around 35 kilobases (kb) throughout the whole genome. The use of chromosomal microarrays has made it easier to identify novel microdeletion disorders and enhance the understanding of those that already exist [27].

Chromosomal microarray/array CGH involves attaching brief DNA sequences that match certain chromosomal locations throughout the whole genome to a solid surface. The arrangement of these sequences impacts the magnitude of the most minute detectable chromosomal abnormality. The standard array will include loci often associated with microdeletion/duplication disorders, along with many subtelomeric and pericentromeric areas. Subtelomeric regions are often affected by DNA copy number alterations, as reported by Gresham et al. in [28]. In order to carry out the test, DNA that has been labeled with fluorescent markers is combined with both the patient's DNA and a control sample. This mixture is then applied to the array. Distinct fluorescent probes are used for the patient and control groups. Following hybridization, the signals are identified, and the data produced is analyzed using software to discover any alterations in the number of copies between the control and patient DNA. Testing is conducted on whole blood samples, and the typical time it takes to complete the process might range from one to three weeks, depending on the laboratory and any further investigations required for confirmation (**Figure 2**).

Typical reasons for doing CGH testing include developmental delay, inability to grow, distinctive physical characteristics, numerous birth defects, below-average height, seizure problems, and autism spectrum disorder. Comparative genomic hybridization (CGH) may also be helpful in identifying the individual genes

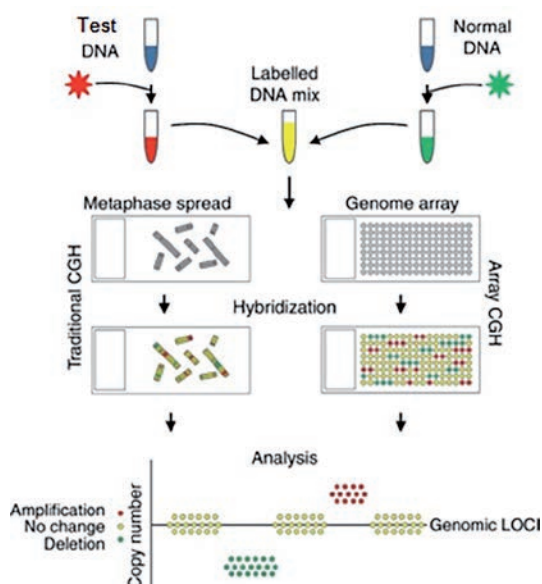


Figure 2.
 Schematic diagram showing the stepwise procedure of CGH and Array CGH (Adopted from Nagarajan et al. [29]).

implicated and in determining the size and nature of a chromosomal aberration found using traditional cytogenetics. According to current practice standards, it is recommended to use CGH in addition to traditional karyotype analysis for patients who have developmental delay or congenital malformations, as stated by Jang et al. in [30]. CGH has limits, including as the possibility of false-negative results if the patient has a copy number alteration that is not detected by the particular array employed in the laboratory. According to Evangelidou et al. [31], point mutations and duplications/deletions that are less than the resolution limit of the specific array (usually 50–100 Kb) cannot be identified [30]. Array CGH is a high throughput method that can be applied to detect copy number changes to a resolution as low as 1 Kb. Genome-wide arrays are rapidly replacing conventional karyotyping in postnatal diagnostics as they are increasingly performed for the evaluation of individuals with birth defects, dysmorphic features, and mental retardation. ISCA (International Standard Cytogenomic Array) Consortium [32] supports the use of array CGH as a first-line test for various phenotypes including suspected microdeletion syndromes [31].

4. Conclusion

Microdeletion syndromes are difficult to diagnose by conventional cytogenetic techniques. The advancements in cytogenetic technology have, however, empowered clinicians as well as researchers to unveil the submicroscopic deletions when it comes to ruling out their genetic causes. Newer cytogenetic tools like FISH, CGH, and array CGH can now be offered as first-line diagnostic tools for the elucidation of microdeletions, copy number variations, and single nucleotide changes. The advent of molecular cytogenetic technology has transformed the process of identifying new microdeletion syndromes besides characterization of the already existing ones.

Author details

Mahrukh Hameed Zargar^{1*}, Tahir Mohiuddin Malla¹, Faisel Ahmed¹, Zafar A. Shah², Jaffar Abass Wani¹ and Asra Kirmani¹

1 Advanced Centre for Human Genetics, Sher-i-Kashmir Institute of Medical Sciences, Srinagar, J&K, India

2 Department of Immunology and Molecular Medicine, Sher-i-Kashmir Institute of Medical Sciences, Srinagar, J&K, India

*Address all correspondence to: mhameedz@gmail.com

IntechOpen

© 2024 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Li M, Pinkel D. Clinical cytogenetics and molecular cytogenetics. *Journal of Zhejiang University. Science. B.* 2006;**7**:162-163
- [2] Watson CT, Tomas MB, Sharp AJ, Mefford HC. The genetics of microdeletion and microduplication syndromes: An update. *Annual Review of Genomics and Human Genetics.* 2014;**15**:215-244
- [3] Vissers LE, Stankiewicz P. Microdeletion and microduplication syndromes. *Genomic Structural Variants: Methods and Protocols.* 2012;**838**:29-75
- [4] Hunter AG. Outcome of the routine assessment of patients with mental retardation in a genetics clinic. *American Journal of Medical Genetics.* 2000;**90**:60-68
- [5] Kirchhoff M, Bisgaard AM, Bryndorf T, Gerdes T. MLPA analysis for a panel of syndromes with mental retardation reveals imbalances in 5.8% of patients with mental retardation and dysmorphic features, including duplications of the Sotos syndrome and Williams-Beuren syndrome regions. *European Journal of Medical Genetics.* 2007;**50**:33-42
- [6] Madon PF, Athalye AS, Sanghavi K, Parikh FR. Microdeletion syndromes detected by FISH—73 positive from 374 cases. *International Journal of Human Genetics.* 2010;**10**(1-3):15-20
- [7] Xu J, Chen Z. Advances in molecular cytogenetics for the evaluation of mental retardation. In: *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* (Vol. 117, No. 1). New York: Wiley Subscription Services, Inc., A Wiley Company; 2003. pp. 15-24
- [8] Peterson JF, Aggarwal N, Smith CA, Gollin SM, Surti U, Rajkovic A, et al. Integration of microarray analysis into the clinical diagnosis of hematological malignancies: How much can we improve cytogenetic testing? *Oncotarget.* 2015;**6**(22):18845
- [9] Kang JU. Overview of cytogenetic technologies. *Korean Journal of Clinical Laboratory Science.* 2018;**50**(4):375-381
- [10] Nowakowska B, Bocian E. Molecular cytogenetic techniques and their application in clinical diagnosis. *Medycyna Wieku Rozwojowego.* 2004;**8**:7-24
- [11] Simons A, Sikkema-Raddatz B, de Leeuw N, Konrad NC, Hastings RJ, Schoumans J. Genome-wide arrays in routine diagnosis of hematological malignancies. *Human Mutation.* 2012;**33**:941-948
- [12] Park JH, Woo JH, Shim SH, Yang SJ, Choi YM, Yang KS, et al. Application of a target array comparative genomic hybridization to prenatal diagnosis. *BMC Medical Genetics.* 2010;**11**:102
- [13] Dave BJ, Sanger WG. Role of cytogenetics and molecular cytogenetics in the diagnosis of genetic imbalances. *Seminars in Pediatric Neurology.* 2007;**14**:2-6
- [14] Vermeesch JR, Fiegler H, de Leeuw N, Szuhai K, Schoumans J, Ciccone R, et al. Guidelines for molecular karyotyping in constitutional genetic diagnosis. *European Journal of Human Genetics.* 2007;**15**:1105-1114
- [15] Wan TS. Cancer cytogenetics: An introduction. *Methods in Molecular Biology.* 2017;**1541**:1-10

- [16] Chennamaneni R, Gundeti S, Konatam ML, Bala S, Kumar A, Srinivas L. Impact of cytogenetics on outcomes in pediatric acutelymphoblastic leukemia. *South Asian J Cancer*. 2018;7:263-266
- [17] Chandran RK, Geetha N, Sakthivel KM, Aswathy CG, Gopinath P, Raj TVA, et al. Genomic amplification of BCR-ABL1 fusion gene and its impact on the disease progression mechanism in patients with chronic myelogenous leukemia. *Gene*. 2018;686:85-91
- [18] Russo A, Degross F. Molecular cytogenetics of the micronucleus: Still surprising. *Mutation Research*. 2018;836(Pt A):36-40
- [19] Iourov IY, Vorsanova SG, Yurov YB. Recent patents on molecularcytogenetics. *Recent Patents on DNA & Gene Sequences*. 2008;2:6-15
- [20] Bejjani BA, Shaffer LG. Clinical utility of contemporary molecularcytogenetics. *Annual Review of Genomics and Human Genetics*. 2008;9:71-86
- [21] Qiu HR, Miao KR, Wang R, Qiao C, Zhang JF, Zhang SJ, et al. The application of fluorescence in situ hybridization in detecting chronic myeloid leukemia. *Zhonghua Yi Xue Yi Chuan Xue ZaZhi*. 2009;26:207-210
- [22] Li MM, Andersson HC. Clinical application of microarray-based molecular cytogenetics: An emerging new era of genomic medicine. *The Journal of Pediatrics*. 2009;155:311-317
- [23] Patrinos GP, Ansorge W, Phillip B. *Danielson Molecular Diagnostics*. 3rd ed. Greece: Academic Press; 2016. pp. 249-267
- [24] Melissa SS, Mirella JM, Ronald JW. Evolving applications of microarray analysis in prenatal diagnosis. *Current Opinion in Obstetrics & Gynecology*. 2011;23:103-108
- [25] Zvara Á, Kitajka K, Faragó N, Puskás LG. Microarray technology. *Acta Biologica Szegediensis*. 2015;59 (Suppl. 1):51-67
- [26] Van den Veyver IB, Beaudet al. Comparative genomic hybridization and prenatal diagnosis. *Current Opinion in Obstetrics and Gynecology*. 2006;18(2):185-191
- [27] Slavotinek AM. Novel microdeletion syndromes detected by chromosome microarrays. *Human Genetics*. 2008;124:1-17
- [28] Gresham D, Dunham MJ, Botstein D. Comparing whole genomes using DNA microarrays. *Nature Reviews Genetics*. 2008;9(4):291-302
- [29] Nagarajan P, Tetzlaff MT, Curry JL, Prieto VG. Use of new techniques in addition to IHC applied to the diagnosis of melanocytic lesions, with emphasis on CGH, FISH, and mass spectrometry. *Actas Dermo-Sifiliográficas*. 2017;108(1):17-30
- [30] Jang W, Kim Y, Han E, Park J, Chae H, Kwon A, et al. Chromosomal microarray analysis as a first-tier clinical diagnostic test in patients with developmental delay/intellectual disability, autism spectrum disorders, and multiple congenital anomalies: A prospective multicenter study in Korea. *Annals of Laboratory Medicine*. 2019;39(3):299
- [31] Evangelidou P, Alexandrou A, Moutafi M, Ioannides M, Antoniou P, Koumbaris G, et al. Implementation of high resolution whole genome array CGH in the prenatal clinical setting: Advantages, challenges, and review

of the literature. Bio Med Research
International. 2013;**2013**(1):346762

[32] Miller DT, Adam MP, Aradhya S,
Biesecker LG, Brothman AR,
Carter NP, et al. Consensus statement:
Chromosomal microarray is a first-tier
clinical diagnostic test for individuals
with developmental disabilities or
congenital anomalies. The American
Journal of Human Genetics.
2010;**86**(5):749-764

Chapter 3

Recent Molecular Techniques in Cytogenetics

*Rikhotso Mpumelelo Casper, Kachienga Leonard,
Magwalivha Mpho and Nethathe Bono*

Abstract

Cytogenetics involves all aspects of chromosome biology. Through the application of molecular cytogenetic techniques in biology and medicine, we gain insights into the structural and functional organization of chromosomes, chromosomal abnormalities, and genomic variations in developmental, medical, and cancer genetics. This chapter explores recent advancements in molecular cytogenetics and their implications for managing hereditary diseases. It explores cutting-edge molecular methodologies such as spectral karyotyping, fluorescence *in situ* hybridization, and next-generation sequencing alongside the groundbreaking CRISPR genome editing technology. The chapter will highlight how these technologies have enhanced our understanding of chromosomal abnormalities, genome variation, and the genetic basis of hereditary diseases. Furthermore, it will shed light on the potential of these advancements in paving the way for novel therapeutic interventions aimed at rectifying genetic abnormalities associated with hereditary diseases, thereby offering new avenues for medical intervention and treatment.

Keywords: molecular cytogenetics, genome editing, next-generation sequencing, chromosomal aberrations, hereditary diseases

1. Introduction

Cytogenetics is a discipline of biology that focuses on the study of chromosomes and their heredity, particularly when applied to medical genetics [1]. The history of cytogenetics began with the description of thread-like structures in plant cell nuclei by Swiss botanist Nageli in the 1840s [2]. These structures are now known as chromosomes. These chromosomes become visible under a microscope during cell division, and specialized staining techniques enable the examination of their number and structure in diagnostic testing [2].

The science of cytogenetics is based on the genetics of cells whereby information from the parent material is passed to the offspring through hereditary [3]. This is determined by the functions and characteristics of the cell transmitted from parents to the offspring. This process applies to the germ cells of the ovary and testes, which, through fertilization become a new offspring and the somatic cells that produce offspring cells of the same tissue or organ where they reside [3]. Li and Pinkel [4] also defined cytogenetics

as the relationship between aberrations (defects) in chromosomes and human genetic diseases. Furthermore, this area of genetics also enables the study of fundamental processes. One of them is the nature of inherited syndromes. These genomic alterations are involved in tumorigenesis, the whole genome of organisms or humans in a three-dimensional organization and epigenetic features of higher-order chromatin structure [4, 5]. Cytogenetics includes all aspects of chromosome biology as well as the use of molecular cytogenetic techniques in biology and medicine, such as chromosome structural and functional organization, genome variation, expression and evolution, chromosome abnormalities, and genomic variations in medical genetics and tumor genetics.

In clinical practice, cytogenetics plays a crucial role in evaluating infants and children with birth defects, investigating individuals experiencing infertility or recurrent pregnancy loss, analyzing prenatal and pre-implantation samples for chromosomal abnormalities, and identifying cancer-related chromosomal alterations in patients with solid tumors and hematologic neoplasms [6]. These same methodologies are also employed in cytogenetic research focusing on speciation, genome stability, and chromosomal evolution.

2. Molecular cytogenetic techniques

Molecular cytogenetic techniques are a vital component of the diagnostic process of genetic disorders caused by chromosomal abnormalities [6]. These techniques have been developed to get beyond the drawbacks of traditional cytogenetics, namely its poor resolution and incapacity to detect specific chromosomal abnormalities. Molecular cytogenetics includes various techniques briefly discussed below such as fluorescence *in situ* hybridization (FISH), multicolor FISH (M-FISH, SKY, and CCK), primed *in situ* labeling (PRINS), next-generation sequencing, spectral karyotyping and comparative genomic hybridization (CGH) [6].

2.1 Spectral karyotyping (SKY)

SKY is a molecular cytogenetic process that utilizes differential visualization techniques of all the chromosomes of human or organisms through image exposure with distinct colors within a single hybridization [7–10]. It is also a process of staining chromosomes to analyze the chromosome complement of an organism, including the structure and number of chromosomes. The chromosome staining is done by the banding technique, which helps to visualize the dark (gene-rich) and light (gene-poor) regions of the chromosome [11]. SKY has been developed to clearly show and identify all 24 human chromosomes at once without requiring prior knowledge of any abnormalities involved. SKY can discern the abnormalities that cannot be detected by traditional banding techniques and fluorescent *in situ* hybridization (FISH). Making SKY a hypersensitive, hyper-accurate, and hyper-intuitionist method [12].

This technique also uses an integration of Fourier spectroscopy with epifluorescence microscopy with charged-coupled device imaging, according to Garini et al. [13], where probes are PCR-labeled through the incorporation of either haptenized or directly labeled nucleotides as a single reaction as dUTP labeled. These specific probe chromosomes are pooled together at the same time, and repetitive sequences are suppressed with excess cot-1 DNA in the hybridization mixture onto metaphase chromosomes. The hybridized chromosomes are then visualized using Fourier spectroscopy with epifluorescence microscopy with charged-coupled device imaging that allows

triple bandpass to filter and permits simultaneous excitation of all the fluorochromes and the entire emission spectrum of metaphase in the range of between 400 and 800 nm [10]. Once a single image containing the spectra information for each image is obtained, the fluorescent intensities in the green, red and near infrared emission range are visualized in a standard display of red, green, and blue (RGB) [7]. The authors further opined that the pixels with the same spectral information are assigned a pseudocolor, which allows for the spectral classification of all chromosomes and the chromosomes are classified and aligned in a karyotype table. Interpretation and comparison of all aberrations are summarized in the karyogram.

It is used to detect inter-chromosomal structural aberrations (translocations and insertions) which are associated with balanced and imbalanced arrangements. This technique can be applied in the identification of cryptic translocation and expounding on the complex aberrations [14–16]. It is also used to identify various markers and ring chromosomes which are not unidentifiable by conventional banding techniques [17–19]. Intrachromosomal alterations resulting in small deletions or duplications and para- or pericentric inversions do not result in changes in chromosome size or spectral signature of the aberrant chromosome and, therefore, cannot be detected by SKY. Very small marker chromosomes or double minute chromosomes cannot always be unambiguously classified.

2.2 Fluorescence *in situ* hybridization (FISH)

This technique was first used to map the human genome and to localize a specific gene on a chromosome. It is further used in the clinical field to detect chromosomal aberrations such as duplication, deletion, balanced/unbalanced translocations, and monosomy/trisomy [11]. FISH is based on the use of chromosome region-specific and fluorescent-labeled DNA probes. FISH refers to the use of labeled nucleic acid sequence probes for the visualization of specific DNA or RNA sequences on mitotic chromosome preparations or in interphase cells [10]. Pinkel et al. [20, 21] developed a technique to visualize chromosomes using fluorescent-labeled probes called fluorescent *in situ* hybridization (FISH). The technique is the most convincing technique for locating the specific DNA sequences, gene mapping, diagnosis of genetic diseases, and genetic aberrations aiding various types of cancers or identification of novel oncogenes. FISH involves the annealing of RNA or DNA probes attached to a fluorescent reporter molecule with a specific target sequence of sample DNA, which can be viewed under fluorescence microscopy. The technique has advanced; through multicolor whole chromosome probe techniques such as spectral karyotyping, multiplex FISH, or an array-based method using comparative genomic hybridization, it is possible to screen the whole genome simultaneously. FISH has been acknowledged as a dependable diagnostic and discovery tool in the fight against genetic diseases and has transformed the field of cytogenetics [22]. The following processes or steps are involved: (a) double-stranded chromosomes, preferably of metaphase, are denatured (on fixed slides) by using formamide and heat, (b) the slide is pooled with fluorescent-labeled probes at 37°C, (c) fluorescent probes anneal at the complementary site of the chromosomes, and (d) visualize the slide under the fluorescence microscope [23].

2.3 Comparative genomic hybridization (CGH)

This process utilizes the reference DNA together with labeled tumor through hybridization on the normal metaphase chromosomes to generate a profile of DNA

copy in tumor genomes along the chromosome length [10]. It is an important global screening that tests the aberrations in the chromosomes within a given tumor genome. According to Dorritie et al. [10], this process allows for the retrospective identification of aberrations within correlations of cytogenetics and prognosis. The most important thing about this process is that it requires less DNA and has proven useful in establishing a solid tumor progression in phenotype/genotypic [24]. The only drawback to this process is that it does not capture any information regarding the nature of chromosomal segments involved in a copy number of alteration (chromosome rearrangement).

Comparative genomic hybridization (CGH) is a unique FISH method (dual probes) used to identify any genetic abnormalities. The basis of the technique is comparing the total genomic DNA of a particular sample (such as tumor DNA) with the total genomic DNA of normal cells. Usually, two distinct fluorescent dyes are used to mark an identical amount of both tumor and normal DNA. This combination is then added and hybridized to a normal lymphocyte metaphase slide [25]. A fluorescent microscope equipped with a CCD camera and an image analysis system is used to view the slides, and evaluation software is used to count copies of genetic materials.

2.4 Primed *in situ* labeling (PRINS)

Primed *in situ* labeling (PRINS) is a versatile and inexpensive technique that has been used traditionally to identify tandemly repeated target sequences in chromosomes and nuclei as an alternative to FISH. The technology has the benefit of using very small probes that can quickly penetrate practically any target and can differentiate among closely similar DNA sequences *in situ*. This unit covers both the standard PRINS and the dideoxy-PRINS variant, which can significantly boost the reaction's sensitivity [26].

3. Recent high-throughput genomic methods

3.1 Array-CGH

In this throughput, metaphase chromosomes are used in the comparative genome hybridization, which is substituted by a high-density DNA microarray sported with artificial human bacterial chromosome (ABC) cloned [27]. They include cDNA sequences/PCR, degenerated primer PCR, or, for better genomic resolution, rolling circle amplification, which is incorporated [28]. Single nucleotide polymorphism (SNP) is used to assist in detecting rare and common genomic arrangements. This technique requires a single hybridization onto one array common genome hybridization process that is pegged on co-hybridization with a test together with a reference DNA [28].

3.2 Next-generation sequencing (NGS)

This technique is also known as second-generation or parallel sequencing. This approach entails arrays of several thousands of sequences in templates in parallel, generating numerous more copies of short reads of DNA per lane [29, 30]. Most of the common platforms used in this process are the following: 454 (Roche Applied Science, Branford, CT), Solexa (Genome Analyzer, Illumina Inc. San Diego, CA), SOLiD (Applied Biosystems, Foster City, CA), Polonator (Dover/Harvard Systems, Salem, NH), and HeliScope (Helicos BioSciences, Cambridge, MA) [28]. The authors

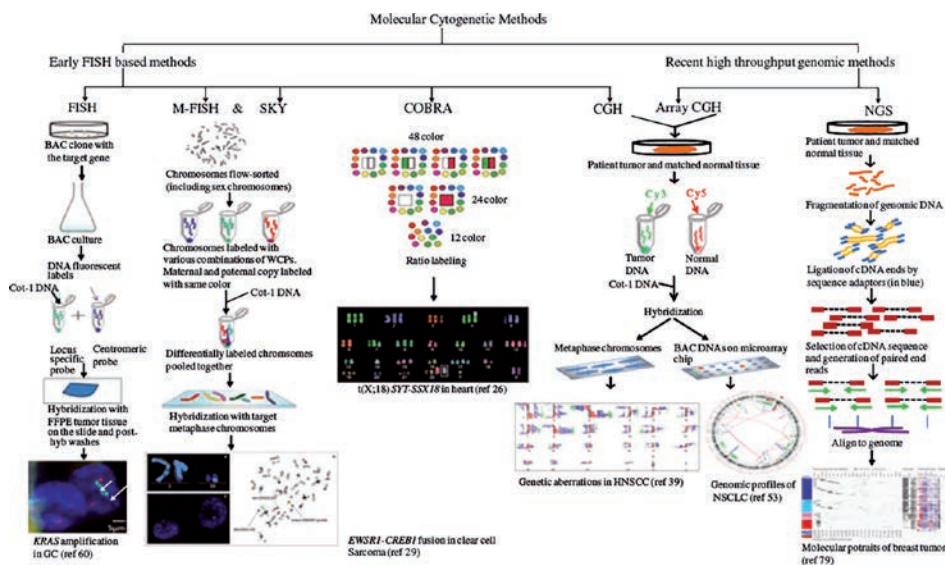


Figure 1.
 Molecular cytogenic techniques and recent high throughput [28].

further reported that the NGS technique has been applied wholesomely in characterizing populations of expressed genes, all protein-coding genes (exomes), as shown in **Figure 1**. NGS also involves fragmentation of the DNA, immobilization of the fragments, and ligation of adapters to form libraries [28]. It is currently used in cancer research to establish disease-causing mutations and drug discovery because of its cost-effective and high sequencing capacity.

4. Applications in chromosomal abnormalities

Molecular cytogenetics plays a vital role in the characterization and detection of chromosomal abnormalities, which are changes in the number or structure of chromosomes that can result in a variety of genetic disorders and diseases. To discover different chromosomal abnormalities, the area of molecular cytogenetics has greatly evolved, including cutting-edge techniques like karyotyping, gene rearrangement investigations, and molecular cytogenetics [31]. There are two types of chromosomal abnormalities: structural and numerical variations. Changes in chromosomal numbers, such as polyploidy (having multiple copies of the entire set of chromosomes) and aneuploidy (having extra or missing chromosomes), are examples of numerical variations. Conversely, structural variants refer to alterations in the chromosomal structure, including translocations, inversions, duplications, Insertion, Ring chromosomes, deletions, and cross-displacement (**Figure 2**).

4.1 The detection and characterization of diseases resulting from chromosomal abnormalities

The detection and characterization of chromosomal abnormalities and the specific role of molecular cytogenetics in understanding structural variations are discussed below:

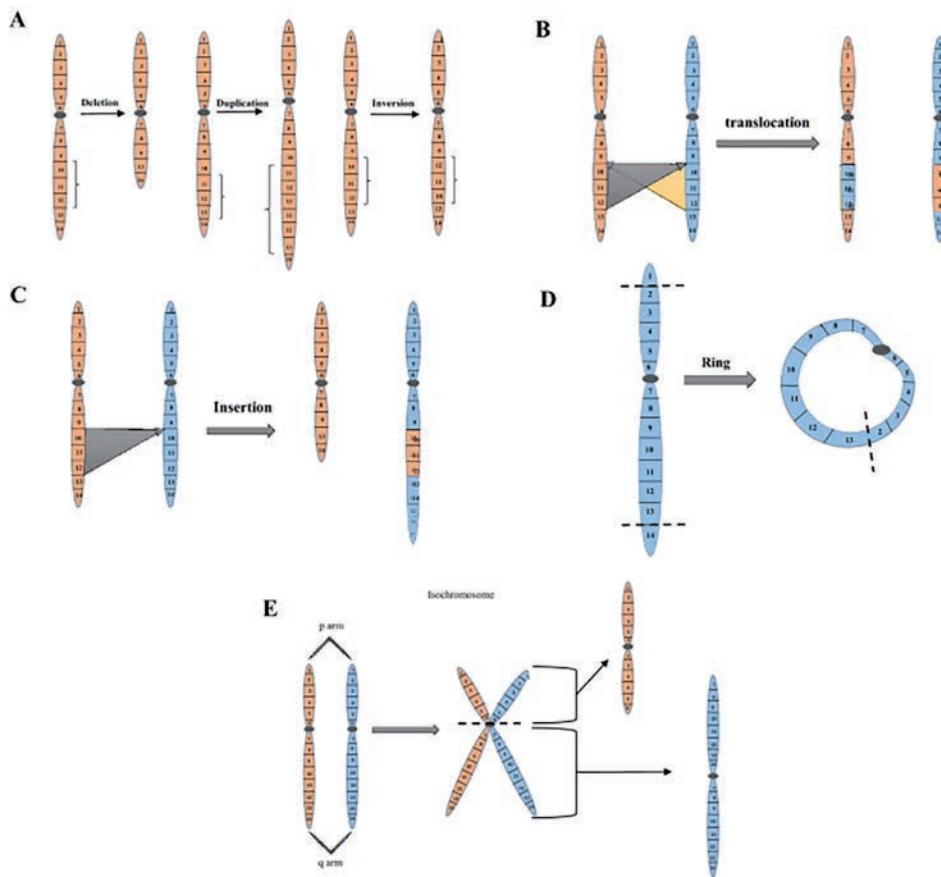


Figure 2. A schematic picture represents the chromosomal abnormalities. Chromosomal abnormalities can be divided into deletion, duplication and inversion, which occur in single chromosome (a), translocation (b) insertion (c) ringing the chromosome (d) and cross-displacement between two chromosomes (e). Picture obtained from Montazerinezhad et al. [31].

4.1.1 Hematologic malignancies

These are oncogenes that have an impact on immune systems such as blood, leukemia, lymph nodes, lymphoma and bone marrow. Combined analysis of FISH-based methods and karyotyping has been useful in the precise delineation of the karyotypes and accurate chromosome analysis of complex diagnostic cases. According to Das and Tan [28], with the introduction of high throughput genomic technologies such as NGS, FISH-based chromosome level detection has gradually changed focus to genome-wide detection of single nucleotide and indel variants that are common in cancer.

4.1.2 Gliomas

These are frequent tumors that occur in the central nervous system, and they are classified based on the cell type or non-neoplastic glial cells (ependymomas, astrocytomas, oligodendrogliomas and mixed gliomas such as oligoastrocytomas) on the supratentorial and infratentorial of the membrane of brain. The initiation

of gliomas is based on multiple genetic alterations. The FISH technique is used to detect oligodendrogliomas which is associated with better response to adjuvant therapy because of imbalance translocation between chromosomes 1 and 19 [27]. The evaluation of the molecular status of each of these markers is a major relevant diagnostic and prognostic tool for brain tumors and is used by neuropathologists in routine clinical practice.

4.1.3 Sarcomas

These oncogenes constitute approximately 1% of all solids tumors, and there is considerable morphological overlap between various histogenetic types. Chromosomal karyotyping is generally the classic approach for identifying genetic markers and translocations in soft tissue sarcomas.

4.1.4 Carcinomas

Carcinomas, too, have highly complex karyotypes like sarcomas and the enumeration of these aberrations by molecular cytogenetics has rapidly increased.

Table 1 summarizes the detection and characterization techniques for the above-mentioned diseases and their advantages and disadvantages.

Method	Techniques	Application	Advantage/disadvantage
Cytogenetics	Karyotype	Detecting numerical and gross structural aberrations	Low resolution, time-consuming, and labor requirements
	FISH	Detecting trisomies, monosomies and microdeletions	Detects mosaicism
	CGH	Detects copy number variations of genetic material	Used only for losses and gains
Molecular	RFLP	Restriction fragments are separated by electrophoresis	Requires mutation in restriction site
	ARMS PCR	Allele-specific amplification of mutant and normal allele, determination of the genotype of an individual	Highly sensitive possible to detect any known mutation May increase time and costs
	Multiplex PCR	Amplification of more than one target simultaneously	Reduces time and labor requirements, lower sensitivity and specificity
	Nested PCR	Amplification using external and internal primer sets	More sensitive, decreases nonspecific amplification
	RT-PCR	Amplification of RNA	Amplification of all RNA types May increase time and costs
	Real-time PCR	Amplification, detection, and quantification of target	Increased specificity Usually eliminates postamplification analyzes More expensive
	MLPA	Deletions and duplications	A multiplex technique Identifies very small single gene aberrations (50–70 nt)
	DGGE	Based on migration within gradient gel electrophoresis	Detects close to 100% of point mutations

Method	Techniques	Application	Advantage/disadvantage
	SSCP	Based on migration within gel electrophoresis	Detects about 80–90% of point mutations
	Heteroduplex analysis	Based on homoduplexes and heteroduplexes motilities in gel electrophoresis	Detects nearly 80% of mutations
	CCM	DNA: DNA or DNA: RNA heteroduplexes are cleaved by piperidine	All possible mutations are detectable, uses toxic substances
	PTT	It is based on a combination of PCR, transcription, and translation	Detects translation-terminating mutations, missense mutations are not detected
	OLA	It is based on the ligation of two flanked primers annealed with target sequences	Detects all base exchanges

FISH: fluorescence in situ hybridization; CGH: comparative genomic hybridization; RFLP: restriction fragment length polymorphism; ARMS: amplification refractory mutation system; PCR: polymerase chain reaction; RT: reverse transcriptase; MLPA: multiplex ligation-dependent probe amplification; DGGE: denaturing gradient gel electrophoresis; SSCP: single strand conformational polymorphism; CCM: chemical cleavage of mismatch; PTT: protein truncation test; OLA: oligonucleotide ligation assay.

Table 1.
Cytogenetics and molecular methods for mutation detection [25].

5. Genome variations and genetic diseases

The molecular cytogenetic techniques have significantly impacted the understanding of genome variations and genetic diseases. These techniques, such as fluorescence *in situ* hybridization (FISH), have enhanced the resolution and diagnostic utility of cytogenetic analysis, enabling the specific detection of unique sequences, chromosomal regions, or entire chromosomes in cells [22, 32]. They have played a crucial role in identifying disease-causing mutations and variants by providing insights into genomic variations in unaffected individuals and those with morbid conditions affecting all cells of an organism [6]. In the context of hereditary diseases, molecular cytogenetics has become a vital tool for diagnosing genetic disorders caused by chromosomal abnormalities. By combining cytogenetics and molecular biology, molecular cytogenetics has increased the resolution of cytogenetic analysis, allowing for the detection of subtle and gross chromosome rearrangements, heteromorphisms, and other genomic variations that can have diverse phenotypic effects [6]. These techniques have revolutionized the field by enabling the study of fundamental biological questions related to inherited syndromes, tumorigenesis, and the three-dimensional organization of the human genome. By utilizing molecular cytogenetic approaches, researchers can identify chromosomal abnormalities with higher precision and delve into the genetic basis of hereditary diseases, ultimately aiding in the identification of disease-causing mutations and variants that underlie various genetic disorders [6].

6. Therapeutic implications

The potential of molecular cytogenetic advancements in therapeutics lies in the ability to correct genetic abnormalities through gene therapy approaches [33, 34]. Gene

therapy involves altering the genetic code to restore the functions of critical proteins affected by genetic mutations. This can be achieved through gene transfer therapy, where a normal copy of the gene is introduced to replace a faulty one, or through genome editing, which modifies existing DNA in cells. The use of viral vectors, such as adenoviruses and lentiviruses, has been successful in delivering genetic material for gene therapy, although it carries some risks like triggering immune responses or causing errors leading to cancer [35–37]. Researchers are also exploring non-viral vector technologies, like nanoparticles, to deliver genetic material more safely and effectively. Challenges in gene therapy include ensuring safety and efficiency, and addressing the risks and benefits specific to each disease being treated [38]. The development of gene therapy requires careful consideration of factors like sustained manufacturing of therapeutic gene products without harmful side effects [37, 39]. Future directions in therapeutic interventions involve overcoming technical challenges, improving the delivery of genes or gene-editing tools to target specific cells, and enhancing control over the functionality and timing of treatments within the body. As gene therapy continues to evolve, advancements in vector technologies, delivery systems, and understanding of genetic diseases will pave the way for more effective and targeted therapeutic interventions.

7. Case studies and clinical applications

The sources provided offer valuable insights into molecular cytogenetics and its clinical applications. Molecular cytogenetics, a field combining molecular biology and cytogenetics, focuses on analyzing chromosome structure to distinguish normal and cancer-causing cells. It plays a crucial role in diagnosing and treating various malignancies, including hematological malignancies and brain tumors. Techniques like fluorescence *in situ* hybridization (FISH) are commonly used to detect chromosomal abnormalities associated with genetic diseases and cancer [6]. In clinical settings, molecular cytogenetics is instrumental in diagnosing genetic disorders caused by chromosomal abnormalities. It has significantly advanced diagnostic capabilities beyond classical cytogenetics by increasing resolution and diagnostic utility. Techniques like FISH, multicolor FISH, and comparative genomic hybridization (CGH) are pivotal in identifying chromosomal abnormalities and understanding their diagnostic implications [40]. Case studies in molecular cytogenetics demonstrate its diagnostic and prognostic applications in oncology and medical genetics. By analyzing chromosome structure, these studies provide insights into the genetic causes of cancer, potential treatment strategies, and personalized medicine approaches based on cytogenetic findings. The field continues to evolve, with ongoing projects like the Cancer Genome Characterization Initiative (CGCI) focusing on genomic research for rare cancers to elucidate genetic abnormalities and their role in cancer pathogenesis [40].

8. Ethical considerations and future directions

The ethical considerations surrounding genome editing and genetic manipulation are a topic of significant debate and concern. Various international bodies and studies have addressed these issues, emphasizing the importance of ethical guidelines, transparency, responsible science, respect for persons, fairness, and transnational cooperation [41]. One key ethical concern is the distinction between

somatic and germline interventions, as well as between therapy and enhancement. The difficulty in drawing clear lines between these categories has led to debates on the normative implications of human nature, consent from future generations, potential slippery slopes toward eugenics, and implications for justice and equality [42]. Regulatory frameworks play a crucial role in governing genome editing technologies. Oversight of human genome editing is embedded within the larger context of international conventions and norms for the protection of human rights and research involving human subjects. Principles for governance of genome editing have general application globally and are reflected in specific statutory and regulatory rules adopted by various nations [41]. Looking toward the future, it is essential to continue developing robust ethical guidelines, fostering public engagement and dialog, establishing global regulatory bodies, and promoting international collaborations to ensure that the use of gene-editing technologies is grounded in international ethical standards and addresses the ethical challenges associated with these advancements [43].

9. Conclusion

In conclusion, molecular cytogenetics continues to be a vital field in both biological and medical research, offering invaluable insights into the intricate world of chromosome biology. Through cutting-edge techniques such as spectral karyotyping, fluorescence *in situ* hybridization, next-generation sequencing, and CRISPR genome editing, researchers have made significant strides in understanding chromosomal abnormalities, genome variations, and the genetic underpinnings of hereditary diseases. The advancements discussed in this chapter not only deepen our comprehension of genetic disorders but also hold immense promise for clinical applications. The potential to correct genetic abnormalities through genome editing opens new avenues for medical treatment, offering hope to individuals affected by previously incurable conditions. The advancements in molecular cytogenetics outlined in this chapter have the potential to revolutionize the way we diagnose, manage, and treat hereditary diseases, heralding a new era of precision medicine.

Author details

Rikhotso Mpumelelo Casper¹, Kachienga Leonard¹, Magwalivha Mpho¹
and Nethathe Bono^{2*}

1 Faculty of Sciences, Department of Biochemistry and Microbiology, Engineering and Agriculture, University of Venda, Thohoyandou, South Africa

2 Faculty of Sciences, Department of Food Science and Technology, Engineering and Agriculture, University of Venda, Thohoyandou, South Africa

*Address all correspondence to: bono.nethathe@univen.ac.za

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Zhang Z, Baxter AE, Ren D, Qin K, Chen Z, Collins SM, et al. Efficient engineering of human and mouse primary cells using peptide-assisted genome editing. *Nature Biotechnology*. 2024;**42**(2):305-315
- [2] Durmaz AA, Karaca E, Demkow U, Toruner G, Schoumans J, Cogulu O. Evolution of genetic techniques: Past, present, and beyond. *BioMed Research International*. 2015;**1**:461524
- [3] Niel Wald MD. Cytogenetics. *Archives of Environmental Health: An International Journal*. 1962;**5**(3):217-228. DOI: 10.1080/00039896.1962.10663274
- [4] Li M, Pinkel D. Clinical cytogenetics and molecular cytogenetics. *Journal of Zhejiang University Science B*. 2006;**7**:162-163
- [5] Speicher MR, Carter NP. The new cytogenetics: Blurring the boundaries with molecular biology. *Nature Reviews Genetics*. 2005;**6**(10):782-792
- [6] Nowakowska B, Bocian E. Molecular cytogenetic techniques and their application in clinical diagnosis. *Medycyna Wieku Rozwojowego*. 2004;**8**(1):7-24
- [7] Liyanage M, Coleman A, Manoir SD, Veldman T, McCormack S, Dickson RB, et al. Multicolour spectral karyotyping of mouse chromosomes. *Nature Genetics*. 1996;**14**(3):312-315
- [8] Knutsen T, Veldman T, Padilla-Nash H, Schröck E, Liyanage M, Ried T. Spectral karyotyping: Chromosomes in colour. *Applied Cytogenetics*. 1997;**23**:26-32
- [9] Macville M, Veldman T, Padilla-Nash H, Wangsa D, O'Brien P, Schröck E, et al. Spectral karyotyping, a 24-colour FISH technique for the identification of chromosomal rearrangements. *Histochemistry and Cell Biology*. 1997;**108**:299-305
- [10] Dorritie K, Montagna C, Difilippantonio MJD, Ried T. Advanced molecular cytogenetics in human and mouse. *Expert Review of Molecular Diagnostics*. 2004;**4**(5):663-676
- [11] Keagle MB, Gersen SL. Basic cytogenetics laboratory procedures. In: *The Principles of Clinical Cytogenetics*. New York, NY: Springer New York; 2012. pp. 53-65
- [12] Guo B, Han X, Wu Z, Da W, Zhu H. Spectral karyotyping: A unique technique for the detection of complex genomic rearrangements in leukemia. *Translational Paediatrics*. 2014;**3**(2):135
- [13] Garini Y, Macville M, du Manoir S, Buckwald RA, Lavi M, Katzir N, et al. Spectral karyotyping. *Bioimaging*. 1996;**4**(2):65-72
- [14] Schröck E, Veldman T, Padilla-Nash H, Ning Y, Spurbeck J, Jalal S, et al. Spectral karyotyping refines cytogenetic diagnostics of constitutional chromosomal abnormalities. *Human Genetics*. 1997;**101**:255-262
- [15] Coleman AE, Schröck E, Weaver Z, Du Manoir S, Yang F, Ferguson-Smith MA, et al. Previously hidden chromosome aberrations in T (12;15)-positive BALB/c plasmacytomas uncovered by multicolor spectral karyotyping. *Cancer Research*. 1997;**57**(20):4585-4592
- [16] Heng HH, Ye CJ, Yang F, Ebrahim S, Liu G, Bremer SW, et al. Analysis of marker or complex chromosomal

rearrangements present in pre-and post-natal karyotypes utilizing a combination of G-banding, spectral karyotyping, and fluorescence in situ hybridization. *Clinical Genetics*. 2003;**63**(5):358-367

[17] Haddad BR, Schröck E, Meck J, Cowan J, Young H, Ferguson-Smith MA, et al. Identification of de novo chromosomal markers and derivatives by spectral karyotyping. *Human Genetics*. 1998;**103**:619-625

[18] Huang B, Ning Y, Lamb AN, Sandlin CJ, Jamehdor M, Ried T, et al. Identification of an unusual marker chromosome by spectral karyotyping. *American Journal of Medical Genetics*. 1998;**80**(4):368-372

[19] Mrózek K, Iliszko M, Ryś J, Babińska M, Niezabitowski A, Bloomfield CD, et al. Spectral karyotyping reveals 17; 22 fusions in a cytogenetically atypical dermatofibrosarcoma protuberans with a large marker chromosome as a sole abnormality. *Genes, Chromosomes and Cancer*. 2001;**31**(2):182-186

[20] Pinkel D, Gray J, Trask B, Van den Engh G, Fuscoe J, Van Dekken H. Cytogenetic analysis by in situ hybridization with fluorescently labeled nucleic acid probes. In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 51. New York, United States: Cold Spring Harbor Laboratory Press; 1986a. pp. 151-157

[21] Pinkel D, Straume T, Gray JW. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences*. 1986b;**83**(9):2934-2938

[22] Shakoory AR. Fluorescence in situ hybridization (FISH) and its applications. In: Bhat T, Wani A, editors.

Chromosome Structure and Aberrations. New Delhi: Springer; 2017. pp. 343-367

[23] Ponnuraj KT. Cytogenetic techniques in diagnosing genetic disorders. In: *Advances in the Study of Genetic Disorders*. Croatia: IntechOpen; 2011. pp. 45-64

[24] Ried T, Knutzen R, Steinbeck R, Blegen H, Schröck E, Heselmeyer K, et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumours. *Genes, Chromosomes and Cancer*. 1996;**15**(4):234-245

[25] Mahdiah N, Rabbani B. An overview of mutation detection methods in genetic disorders. *Iranian Journal of Paediatrics*. 2013;**23**(4):375

[26] Koch J. Principles and applications of PRINS in cytogenetics. *Current Protocols in Human Genetics*. 2004;**42**(1):4-12

[27] Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genetics*. 1998;**20**(2):207-211

[28] Das K, Tan P. Molecular cytogenetics: Recent developments and applications in cancer. *Clinical Genetics*. 2013;**84**(4):315-325

[29] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;**437**(7057):376-380

[30] Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an

evolved bacterial genome. *Science*. 2005;**309**(5741):1728-1732

[31] Montazerinezhad S, Emamjomeh A, Hajieghrari B. Chromosomal abnormality, laboratory techniques, tools and databases in molecular cytogenetics. *Molecular Biology Reports*. 2020;**47**(11):9055-9073

[32] Cui C, Shu W, Li P. Fluorescence in situ hybridization: Cell-based genetic diagnostic and research applications. *Frontiers in Cell and Developmental Biology*. 2016;**4**:89

[33] Shahryari A, Saghaeian Jazi M, Mohammadi S, Razavi Nikoo H, Nazari Z, Hosseini ES, et al. Development and clinical translation of approved gene therapy products for genetic disorders. *Frontiers in Genetics*. 2019;**10**:868

[34] Sheikh-Hosseini M, Larijani B, Gholipour Kakroodi Z, Shokoohi M, Moarefzadeh M, Sayahpour FA, et al. Gene therapy as an emerging therapeutic approach to breast cancer: New developments and challenges. *Human Gene Therapy*. 2021;**32**(21-22):1330-1345

[35] Amer MH. Gene therapy for cancer: Present status and future perspective. *Molecular and Cellular Therapies*. 2014;**2**:1-19

[36] Kotterman MA, Chalberg TW, Schaffer DV. Viral vectors for gene therapy: Translational and clinical outlook. *Annual Review of Biomedical Engineering*. 2015;**17**:63-89

[37] Shirley JL, de Jong YP, Terhorst C, Herzog RW. Immune responses to viral gene therapy vectors. *Molecular Therapy*. 2020;**28**(3):709-722

[38] Dunbar CE, High KA, Joung JK, Kohn DB, Ozawa K, Sadelain M. Gene therapy comes of age. *Science*. 2018;**359**(6372):4672

[39] Goswami R, Subramanian G, Silayeva L, Newkirk I, Doctor D, Chawla K, et al. Gene therapy leaves a vicious cycle. *Frontiers in Oncology*. 2019;**9**:297

[40] Liehr T, Claussen U. Current developments in human molecular cytogenetic techniques. *Current Molecular Medicine (Sharjah, Sharjah, United Arab Emirates: Bentham Science Publishers)*. 2002;**2**(3):283-297

[41] National Academies of Sciences, National Academy of Medicine, National Academy of Sciences, Committee on Human Gene Editing, Scientific, Medical and Ethical Considerations. *Human Genome Editing: Science, Ethics, and Governance*. National Academies Press; 2017

[42] Almeida M, Ranisch R. Beyond safety: Mapping the ethical debate on heritable genome editing interventions. *Humanities and Social Sciences Communications*. 2022;**9**(1):1-14

[43] Ghosh S, Ghosh S, Raghunath M, Bhaskar R, Sinha JK. Balancing potential benefits and ethical considerations of gene editing. *The Lancet*. 2023;**401**(10394):2109-2110

Technological Advances in Cancer Cytogenetics

Subrata Kumar Dey

Abstract

The field of cytogenetics has witnessed significant changes due to constant evolution of new methodologies for the detection of chromosomal alterations in number and structure. The discovery of Philadelphia chromosome in Chronic Myelogenous Leukemia (CML) ushered in new era in cancer cytogenetics. Classical cytogenetics is based on G-banded chromosome analysis and karyotyping, which can detect alterations in structure and number of chromosomes in a diploid set. Subsequent development of molecular cytogenetic techniques such as Fluorescence In situ Hybridization (FISH) where fluorescent dye-labelled DNA probes are used to identify chromosome rearrangements at metaphase or interphase stage. On the other hand, Comparative Genomic Hybridization (CGH) technique is applied for the detection of copy number variations. It is used to detect chromosomal aberrations such as deletions, additions and ploidy abnormalities, whereas array CGH compares two differentially fluorescent labelled test and reference samples simultaneously at multiple loci. The major advantage of this technique is that it can detect deletions and duplications at submicroscopic level. Furthermore, genome sequencing technology has also been applied to understand the mutational signatures in the origin of different types of cancer. The present review traces the evolution of cytogenetic techniques in the detection of genomic alterations in cancer.

Keywords: cancer, genome, cytogenetics, mutation, techniques

1. Introduction

The determination of diploid chromosome number in human [1] soon led to the discovery of several chromosomal disorders. Over the years, conventional karyotyping was recognized as gold standard for the detection of chromosomal abnormalities. The association of Philadelphia chromosome (translocation between chromosomes 9 and 22) in chronic myelogenous leukemia (CML) [2], specifically relates chromosome abnormality with cancer. In case of acute lymphoblastic leukemia (ALL), translocation was detected between chromosomes 12 and 21 [3] by conventional karyotyping. However, to overcome the limitation of conventional karyotyping and for rapid detection of chromosomal abnormalities, FISH (fluorescence *in situ* hybridization) was developed as an indispensable method for the detection of chromosomal rearrangements such as translocation, inversion, and microdeletion, which has greatly enhanced the efficiency and accuracy of the chromosome analysis [4, 5] by

combining conventional cytogenetics and molecular technologies. On the other hand, analysis of copy number variations and genome-wide screening analysis was made possible by the introduction of CGH (comparative genomic hybridization) technique [6]. This technique was developed to compare the chromosome complements of the solid tumor with that of normal tissue, and its resolution was at the level of five to ten megabases. This method was found to be more accurate than conventional Giemsa banding and FISH. However, the low resolution, which was the main limitation of the CGH method, was further overcome by the introduction of array CGH where metaphase chromosome was replaced by cloned DNA fragments. This facilitates the detection of structural variations of genome at base pair level [7]. Advances in our knowledge of cancer diagnosis and management have led to the introduction of genomic profiling of cancer patients by introducing NGS (next-generation sequencing) technology, which has been found to be very useful in guiding the treatments of cancer patients [8]. Moreover, cancer is characterized by intertumoral and intratumoral variations of cells, and there is an irregular distribution of genetically different tumor subpopulations within and between single disease site or across different disease sites [9]. Progress in high-throughput technologies has enabled a better understanding of genomic alterations in human cancer, and next-generation sequencing is now advised to guide therapy of different types of cancers since the high-throughput technology has played an important role in understanding altered genetic pathways in cancer and to guide targeted therapy [10].

In this review, discussion has been made regarding the technological advancements from conventional cytogenetics to cytogenomic analysis of cancer and clinical usefulness of these techniques in the diagnosis and management of different types of malignancies.

2. Conventional cytogenetic analysis: Karyotyping and chromosome banding

The chromosome constitution of a diploid individual is described as karyotype when chromosomes are arranged on the basis of decreasing order of length and position of centromere. Metaphase chromosome preparation involves short-term cell culture followed by colchicine pretreatment to arrest dividing cells at metaphase stage. Chromosome preparation is done using the conventional air-drying technique, and metaphase plates are subjected to G-banding, which involves trypsin pre-digestion followed by staining with Giemsa stain. As a result, dark bands are termed as G-bands, while light bands are called R-bands. The construction of karyotype from metaphase plates provides an overall picture of the genome of an affected cell. Construction of G-banded karyotype and analysis under microscope enables the identification of different types of structural chromosomal rearrangements such as translocation between chromosome numbers 9 and 22 in acute lymphoblastic leukemia (ALL) [11].

Moreover, alteration in the number or structure of the chromosome, including deletions, duplications, and translocations in the range of 5–10 megabases (Mb), could be detected, which is visible under microscope. One of the advantages of this technique is the detection of the abnormalities that are present at very low number, particularly in myeloid neoplasms [12]. It has been observed that karyotype of cancer cells is often altered due to change in the chromosome number which has been observed in most of the carcinoma cells. It has been suggested that aneuploidy is an

essential component of tumorigenesis and changes in chromosome number are seen in around 85% of carcinoma cells [13]. Moreover, chromosomal analysis of malignant cells faces technical challenge as the results are unpredictable and inconsistent. Low mitotic index and poor chromosome morphology are some of the limitations of conventional cytogenetic method.

2.1 Molecular cytogenetics

It involves the use of different methods where DNA probes are labeled with colored fluorescent tags to observe the specific regions of the chromosomes. These techniques may be used to detect chromosomal alterations in dividing as well as nondividing cell populations.

2.2 Fluorescence *in situ* hybridization (FISH)

The standard FISH protocol involves the use of fluorescent-labeled DNA probes that are capable of hybridizing complementary DNA sequences in chromosomal regions within the nucleus of intact cells or in metaphase chromosomes and produce a fluorescent signal which indicates the chromosomal location of a particular gene or DNA sequence and can be observed under a fluorescence microscope with appropriate filters [14, 15] while in CISH (chromogenic *in situ* hybridization) probe is labeled with biotin or digoxigenin. Sensitivity of these techniques is found to vary with the detection of different types of oncogene. CISH is more sensitive in the detection of HER-2/neu oncogene amplification [16]. The first gene to be mapped using FISH was thyroglobulin [17]. FISH provides a higher level of resolution than the standard G-banding technique and facilitates karyotype analysis of nuclei of nondividing cells, providing rapid diagnosis of different types of leukemia such as BCR: ABL1 fusion gene in chronic myelogenous leukemia (CML). This technique was used for the detection of translocation in large cell lymphoma [18], acute myeloid leukemia (AML) [19], and also for the analysis of chromosomal abnormalities in tumors in children [20]. FISH opened the way for the development of a new technology called spectral karyotyping (SKY) or multicolor FISH (M-FISH). This technique allows the labeling of chromosomes with different painting probes at the same time, resulting in the detection of translocations and other complex karyotypes, which are the characteristic features of solid tumors [21]. There are also two main locus-specific Identifier FISH probe systems for the detection of gene rearrangements in cancer investigation: dual color translocation probes and dual color break-apart probe [22]. The dual color translocation probes are designed to detect known partner genes such as BCR-ABL1 that result from translocation between chromosomes 9 and 22 in chronic myelogenous leukemia (CML). On the other hand, dual color break-apart probes are useful for detecting translocation involving multiple translocation partners in mixed lineage leukemia (MLL) where the MLL gene was affected by translocation at 11q23 [22]. FISH techniques no doubt improved the diagnosis of hematologic malignancies as well as solid tumors in comparison to conventional G-banded techniques. Breast cancer is one of the most common type of tumor reported in women, and FISH technique is used to detect the HER2 gene locus. FISH probes are also available for the detection of ALK, ROS1 oncogenes in lung cancer, and CCNE1 in ovarian cancer [16]. Therefore, the applications of advanced FISH techniques facilitate the identification of chromosomal aberrations that are responsible for oncogene activation, which remain unresolved by the application of conventional karyotyping.

2.3 Comparative genomic hybridization (CGH)

This method involves analyzing copy number variations (CNVs) in the DNA of test sample compared to a reference sample. The aim of this technique is to quickly compare two genomic DNA samples arising from two sources such as solid tumor and normal tissue. In CGH technique, DNA from sample in study and the control DNA extracted from an individual with a normal karyotype are differentially labeled with green and red fluorochromes, respectively. Using fluorescence microscope and computer software, the differentially colored fluorescent signals are compared to detect the gain or loss of material in the test sample [7, 23]. CGH is able to detect unbalanced chromosomal abnormalities. However, it is unable to detect balanced chromosomal abnormalities such as translocations, inversions, or ring chromosome that do not affect copy number variations. CGH technology was used to detect the gene encoding the catalytic subunit of phosphatidylinositol 3-kinase (PIK3CA) which is amplified in ovarian cancer [24], thus identifying PIK3CA as an oncogene associated with ovarian cancer. A distinct advantage of this technique is that formalin-fixed and paraffin-embedded tissues can also be used for investigation which made this technique useful for cancer research. However, the resolution of the conventional CGH is a major limiting factor in its clinical applications. Aberrations smaller than 5-10 Mb cannot be detected using conventional CGH [25].

2.4 Array CGH

CGH methods are found to be labor intensive and require the use of metaphase chromosomes, which results in limited resolution. However, development of array-based CGH method does not require the use of metaphase chromosome. This method uses DNA arrays of human genome adhered to microchip. The main advantage of Array CGH is its ability to detect different types of chromosomal abnormalities such as deletions, duplications and amplification of any locus [7, 9]. Array CGH has advanced the conventional cytogenetics from microscope to computer-based analysis, which is able to identify genomic abnormalities at higher resolution. Array CGH enables analysis of whole genome in a single experiment, and until now, its applications have been mainly directed at detecting genomic abnormalities in cancer. This technique is also used to identify multiple forms of chromosomal instability in colorectal cancer [26], genomic copy number alterations in bladder cancer [27] and cervical cancer [28]. This technique is also applied extensively for deletion mapping and identification of tumor suppressor genes. It has also made a significant contribution to understanding tumor biology and increasing resolution of arrays, made possible analysis of genome up to the single base pair level which indicates its potential to identify many genes within this range [29]. Continuous refinement of diagnostic technologies and the subsequent development of DNA microarrays technologies have made possible whole genome analysis not only at high resolution but also open up the possibilities for clinical application. This new technology has opened up the possibility to combine with other genomic profiling technologies such as Single Nucleotide Polymorphism (SNP) and next-generation sequencing (NGS). These techniques grouped together as cytogenomics technologies, leading toward personalized cancer genomics and help to identify specific cancer signatures for the diagnosis and prognosis of cancer [9].

Table 1 shows comparison between conventional cytogenetic and molecular techniques.

Techniques	Approach	Features	Application
G-Banding	Metaphase Chromosome	Transverse bands across each chromosome	Identification of chromosomal aberrations
FISH	Interphase & Metaphase Chr.	Fluorescence dye Labeled DNA as probe	Detection of Mutation
Spectral Karyotyping	Metaphase Chromosome	Fluorescence dye labeled DNA as probe for each chromosome of karyotype	Genome wide screening
CGH	Fixed biopsies	Labeled DNA as probe	Detection of DNA copy no. variation
Array-CGH	Array DNA	Labeled DNA as probe	Detection of chr. aberrations at submicroscopic level
Genome Sequencing	DNA	Sequencing of whole Genome	Identification of mutation

Abbreviations: FISH: Fluorescence In situ hybridization, CGH: Comparative Genomic hybridization.

Table 1.
 Comparison of conventional cytogenetic and molecular techniques.

2.5 Next-generation sequencing (NGS)

Introduction of NGS platform has opened a new vista in genomic profiling of cancer patients. This technology uses parallel sequencing of multiple small fragments of DNA to determine sequence and this high -throughput technology has allowed a rapid increase in the speed as well as decrease in the cost at which a genome of the patient can be sequenced. NGS approaches include both DNA and RNA analysis. It includes targeted sequencing of genes of interest and RNA sequencing facilitates the detection of posttranscriptional modifications, gene fusion and single nucleotide polymorphisms [30]. Lately, introduction of NGS technology has revolutionizing cancer molecular diagnostics due to its many advantages compared to conventional diagnostic methods. Application of this technology has played an important role in understanding altered genetic pathways in cancer, genomic aberrations, insertions and deletions and copy number variations [30]. This technology is mainly used in solid tumor and hematologic cancer diagnosis. Among males, prostate cancer is the leading cause of death in many countries. High tumor heterogeneity is a challenge for formulating treatment for this cancer. Recent advances in NGS technology facilitated the understanding of genetic profile of prostate cancer and two major signaling pathways such as P13K-PTEN-Akt and RTK-Ras-MAPK [31], which subsequently improved the diagnosis and treatment protocols of prostate cancer. Lung cancer is another leading cause of cancer-related death and is often detected at an advanced stage. Application of NGS has identified EGFR/KRAS/ALK alterations which is responsible for lung cancer [32]. On the other hand, the application of NGS technology not only detects BCR-ABL translocation in chronic myelogenous leukemia (CML) but also facilitates the precise identification of novel mutations that may be of great prognostic value and clinical importance [33]. NGS has been successfully employed to identify

novel mutations in a variety of cancers, including bladder cancer, renal cell carcinoma, small cell lung carcinoma, prostate cancer and different types of leukemias [34]. Advancement of NGS technology has made possible the comprehensive molecular profiles of normal and tumor genomes for designing personalized treatment regimens, though the clinical utility and efficacy of comprehensive genomic profiling are still under evaluation [35]. Furthermore, the availability of cancer sequencing data leads to the identification of mutational signatures and carcinogens responsible for specific signature at a single gene level. Along with DNA alterations, several forms of RNA alterations, including gene silencing, have also been reported in human cancer. The computational approach, which includes the application of machine learning and artificial intelligence to classify and predict models based on genomic datasets, is achieving promising results, suggesting a potential diagnostic use, and it will play an important role in the clinical decision-making process [36].

3. Conclusion

Cytogenetic studies have made possible the discovery of the chromosomal basis of cancer through the identification of chromosomal abnormalities like the Philadelphia Chromosome in chronic myeloid leukemia, t (9;22). This technique identifies chromosomal aberrations in the range of 5–10 megabase (Mb). The initial expansion of cancer cytogenetics techniques was with fluorescence *in situ* hybridization (FISH) to assess the genomic alterations more accurately at submicroscopic level (50-100 kb) by labeling DNA probes with a fluorescent dye or, in certain cases, chromogenic *in situ* hybridization (CISH) by labeling with biotin or digoxigenin. Rapid advancement of molecular cytogenetics has made possible the introduction of comparative genomic hybridization (CGH), which has made possible the genome-wide scan to identify copy number variations and to identify chromosomal regions associated with loss or gain of genetic information at the resolution of 5–10 Mb. Keeping in mind, the accuracy and rapidity of the experiment, further advancement was made in array CGH to identify genomic alterations. By using DNA microarrays, it is possible to identify cause of several chromosomal abnormalities in cancer at genomic level with resolution up to 30 kb. Ongoing research led to the development of next-generation sequencing (NGS) technology, which automated the profiling of genome of cancer patients, identified the altered genetic pathways of oncogenes, which led toward personalized treatment of cancer and detection genomic imbalances in the range of 50 kilobase (kb) to 1 megabase (Mb) resolution. Recent applications of machine learning and artificial intelligence in the analysis of genomic data and creating predictor models will play an important role in diagnosis and clinical decision-making process in cancer-affected patients.

Author details

Subrata Kumar Dey
Swami Vivekananda University, Barrackpore, West Bengal, India

*Address all correspondence to: subratadey184@gmail.com

IntechOpen

© 2024 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Tjio JG, Levan A. The chromosome number of man. *Hereditas*. 1956;**42**:1-6
- [2] Nowel PC, Hungerford DA. A minute chromosome in human chronic granulocytic leukemia. *Science*. 1960;**142**:1497
- [3] Johansson B, Merten F, Mitelman F. Clinical and biological abnormalities in childhood and adult acute lymphoblastic leukemia. *Annals of Medicine*. 2004;**36**(7):492-503
- [4] Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. 1973;**243**:290-293
- [5] Wan TS. The role of FISH in hematologic cancer. *International Journal of Hematologic Oncology*. 2012;**1**:71-86
- [6] Kallioniemi A, Kallioniemi OP, Sudar DA, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*. 1992;**258**(5083):818-821
- [7] Pinkel D, Albertson DG. Array comparative genomic hybridization and its application in cancer. *Nature Genetics*. 2005;**37**:11-17. DOI: 10.1038/ng1569
- [8] Coloman R, Miranda J, Nuria RL, Javier H, Lucia GC, et al. Usefulness and real-world outcomes of next generation sequencing testing in patients with cancer: An observational study on the impact of selection based on clinical judgement. *Clinical Medicine*. 2023;**60**:1-14. DOI: 101016/j.eclinm.2023102029
- [9] Ribeiro IP, Melo JB, Isabel MC. Cytogenetics and cytogenomics evaluation in cancer. *International Journal of Molecular Sciences*. 2019;**20**(19):4711
- [10] Tung N, Battelli C, Allen B, et al. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. *Cancer*. 2015;**121**(1):25-33
- [11] Wan TS. Cancer cytogenetics: Methodology revisited. *Annals of Laboratory Medicine*. 2014;**34**(6):413-425. DOI: 10.3343/alm.2014.34.6.413
- [12] Balciuniene J, Ning Y, Lazarus HM, Aikawa V, et al. Cancer cytogenetics in a genomics world: Wedding the old with the new. *Blood Reviews*. 2024;**66**:101209
- [13] Weinberg RA. *The Biology of Cancer*. 2nd ed. NCBI, USA: Garland Science, Taylor & Francis; 2014
- [14] Pinkel D, Gary JW, Trask B, van den Engh G, et al. Cytogenetic analysis by in situ hybridization with fluorescently labeled nucleic acid probes. *Cold Spring Harbor Symposia on Quantitative Biology*. 1986a;**51**:151-157
- [15] Pinkel D, Straume T, Gray JW. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proceedings of the National Academy of Sciences of the United States of America*. 1986b;**83**:2934-2938
- [16] Natalia MC, Kowalewski J, Lewandowska MA. Use of fluorescence in situ hybridization (FISH) in diagnosis and tailored therapies in solid tumors. *Molecules*. 2020;**25**(8):1864
- [17] Landegent JE. Chromosomal localization of a unique gene by

non-autoradiographic in situ hybridization. *Nature*. 1985;**317**:175-177

[18] Johnson WM, Leek J, Swinbank K, et al. The use of fluorescent in situ hybridization for detection of the t (2;5) (p23; q35) translocations in anaplastic large cell lymphoma. *Annals of Oncology*. 1997;**8**:S65-S69

[19] Dodge W, Cruz J, Zamkoff K, Hurd D, et al. Use of fluorescence in situ hybridization to detect minimal residual disease in hematopoietic stem cell assays from peripheral blood stem cells of 2 patients with trisomy 8 acute myeloid leukemia. *Stem Cells and Development*. 2004;**13**(1):23-26

[20] Raimondi SC. Fluorescence in situ hybridization: Molecular probes for diagnosis of pediatric neoplastic diseases. *Cancer Investigation*. 2000;**18**:135-147

[21] Kearney L. Multiplex-FISH(M-FISH): technique, developments and applications. *Cytogenetic and Genome Research*. 2006;**114**:189-198

[22] Wan TS. Cancer cytogenetics: Methodology revisited. *Annals of Laboratory Medicine*. 2014;**34**(6):413-425. DOI: 10.3343/alm.34.6.413

[23] Weiss M, Hermsen M, Meijer G, Van GN, et al. Comparative genomic hybridization. *Molecular Pathology*. 1999;**52**:243-251

[24] Shayesteh L. PIK3CA is implicated as an oncogene in ovarian cancer. *Nature Genetics*. 1999;**21**:99-102. DOI: 10.1038/ng0496-368

[25] Chen J, Zhang D, Yan W, Yang D, Shen B. Genome screening by comparative genomic hybridization. *Trends in Genetics*. 1997;**13**(10):405-409

[26] Gassenbeek M, Howarth K, Rowan AJ, Gorman PA. Combined array-comparative genomic hybridization and single nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple form of chromosomal instability in colorectal cancer. *Cancer Research*. 2006;**66**(7):3471-3479. DOI: 10.1158/0008-5472.CAN-05-3285

[27] Hurst CD, Fiegler H, Carr P, William S, et al. High resolution analysis of genomic copy number alteration in bladder cancer by microarray based comparative genomic hybridization. *Oncogene*. 2004;**23**(12):2250-2263. DOI: 1038/sj.one.1207260

[28] Kloth JN, Oosting J, Van Wezel T, Szuhai K, et al. Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer. *BMC Genomics*. 2007;**20**(8):53. DOI: 10.1186/1471-2164-8-53

[29] Michels E, De Preter K, Roy NV, Speleman F. Detection of DNA copy number alterations in cancer by array comparative genomic hybridization. *Genetics in Medicine*. 2007;**9**:574-584

[30] Serrati S, Summa SD, Pilato B, Petriella D, et al. Next -generation sequencing: Advances and applications in cancer diagnosis. *Oncotargets and Therapy*. 2016;**9**:7355-7365. DOI: 10.2147/OTT.S99807

[31] Chen J, Zhang D, Yan W, Yang D, Shen B. Translational bioinformatics for diagnostic and prognostic prediction of prostate cancer in the next-generation sequencing era. *BioMed Research International*. 2013;**2013**:901578

[32] Lim SM, Kim EY, Kim HR, et al. Genomic profiling of lung

adenocarcinoma patients reveals therapeutic targets and confers clinical benefit when standard molecular testing is negative. *Oncotarget*. 2016;7(17):24172-24178

[33] Black JS, Salto TM, Millis KI, Catherwood MA. The impact of next generation sequencing technologies on haematological research: A review. *Pathogenesis*. 2015;2(1-2):9-16

[34] Guan Y-F, Li GR, Wang RJW, Yu-Ting Y, et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chinese Journal of Cancer*. 2012;31(10):463-470. DOI: 10.5732/cjc.012.10216

[35] Mardis ER. Applying next-generation sequencing to pancreatic cancer treatment. *Nature Reviews. Gastroenterology & Hepatology*. 2012;9:477-486

[36] F-M MA, Blancas S, Herrera-Montalvo LA, Miranda AH. Cancer genomics. *Archives of Medical Research*. 2022;53:8,723-8,773

Section 2

Methods for Understanding Genetic Causality

Discovering Causality with Mendelian Randomization: What It Is and How to Do It

Hayrettin Okut and Rosey Zackula

Abstract

Mendelian Randomization (MR) is a powerful analytical method that assesses causal relationships. When direct exposure is neither feasible nor ethical, this method incorporates an instrumental variable (such as a genetic variant) as a substitute for an exposure, and measures the causative influence directly related to an outcome (e.g., disease). In this way, MR eliminates the need for an unattainable randomized experiment. However, the instrumental variable must meet three core assumptions: be associated with the exposure of interest, unrelated to any confounding factors that impacts the outcome, and affect the outcome only through the exposure. Conducting an MR study requires six basic steps: (1) define the causal association, (2) select an instrumental variable, (3) evaluate core assumptions, (4) conduct an analysis that produces a model, (5) validate the model, and (6) report findings. Various software offers MR analyses: R, Python, SAS, STATA, and other specialized programs. Although MR is a powerful tool, it has challenges and limitations. For example, instrumental variables can influence multiple traits (pleiotropy), be weak instruments, or can be underpowered. Future advancements aim to correct pleiotropy and expand the range of instrumental variables including genetic variants. MR represents a significant advancement in causal inference in medicine; it overcomes the limitations imposed on randomized clinical trials and is superior to observational studies that only measure associations. This chapter will outline current MR methodologies and provide a framework for those interested in investigating causal relationships using genetic variants.

Keywords: Mendelian randomization (MR), causal inference, instrumental variable, statistical methods, genetic epidemiology

1. Introduction

Mendelian Randomization (MR) is an analytical approach within epidemiology and statistical genetics that uses genetic variants -referred to as instrumental variables- to act as substitutes or proxies for adjustable exposures [1]. Its purpose is to evaluate potential causal connections between exposures and particular outcomes. Prior to the inception of MR, the gold standard approach to show causality was randomized trial designs. Yet they can be expensive and may challenge ethical and

practical constraints; for example, demonstrating that cancer is caused by exposure to smoking. Thus, researchers depended predominantly on observational studies and conventional statistical techniques to explore connections between risk factors and outcomes. Although, observational studies establish associations, rather than causation, and are susceptible to biases, such as confounding and reverse causation, making it challenging to determine the true cause-effect relationships between exposures and outcomes. These limitations raise uncertainties about the validity of findings and hinder the ability to establish definitive causal relationships. MR, however, fills a crucial gap between randomized trials and observational studies to discover causal relationships [2–4].

MR can be considered a form of inherent randomized controlled trial (RCT) where genetic allocation follows Mendel's laws of equal segregation and independent assortment [5]. These Laws, first described by Gregor Mendel in 1865, describe how genetic variations are passed from parents to their offspring in a random fashion. The Law of Equal Segregation states that, when two alleles separate during gamete formation, each resulting gamete has an equal probability of containing either allele. The Law of Independent Assortment states that each allele is independent, and thus offspring may inherit any combination of alleles from their parents. Thus, genetic variants mimic both assumptions of randomness: equal probability and independence. **Figure 1A** and **B** briefly summarizes the history of MR, from Mendel's Laws to the present [6].

MR was first introduced as a study design by Katan [7] to uncover causal connections between exposures and outcomes. This novel approach would go on to provide important insights within fields such as medicine, pharmacology, public health, econometric, and social sciences. Gray and Wheatley elaborated upon MR by obtaining unbiased estimations of the effects of a suspected causal factor, while avoiding the requirement for a traditional RCT [8].

In MR, the genetic variants, or single nucleotide polymorphisms (SNPs), act as instrumental variables, facilitating the investigation of whether an exposure of interest (e.g., BMI, fasting glucose, smoking, alcohol use status) causally influences an

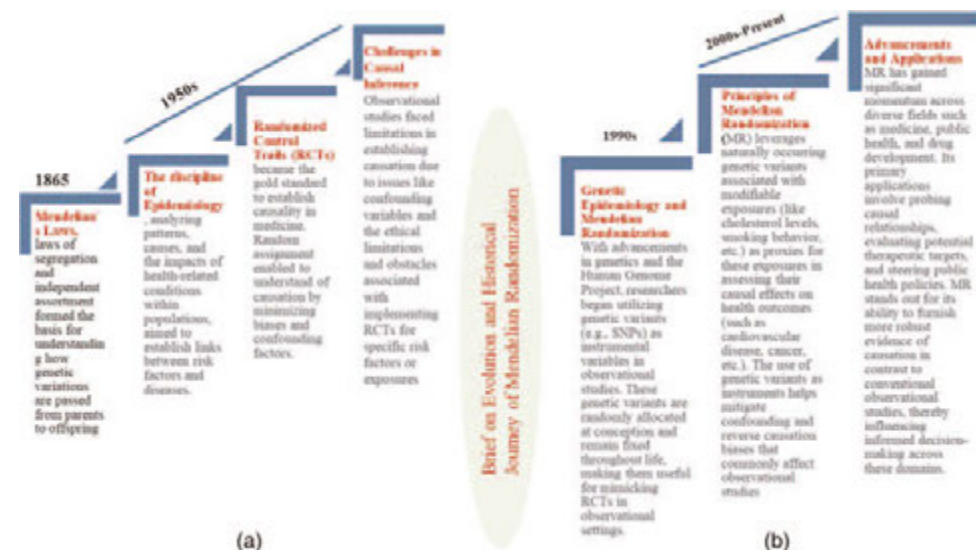


Figure 1.
A. Mendelian Laws; evolution. B. History of Mendelian Randomization (MR).

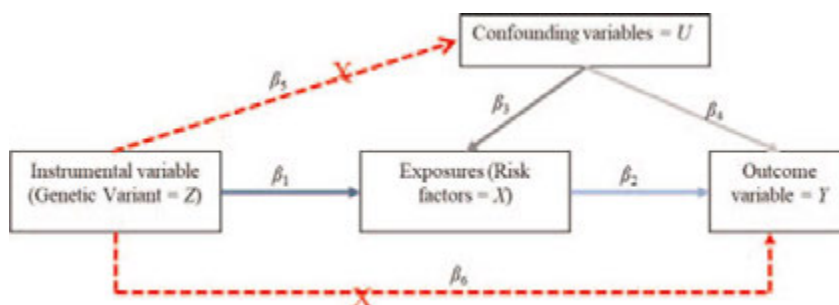


Figure 2. Relationship within Mendelian Randomization (MR). --X-- The red dashed arrows indicate pathways that should not exist if the genetic variant (GV) is valid. The line for β_5 means independent of any unobserved confounding variables, also referred to as the independence assumption ($Z \perp U$). The line for β_6 means GV and outcome Y are independent, conditional on exposure X and unmeasured confounders U ($Z \perp Y|X$), also referred to as the exogeneity assumption. $\rightarrow$ The dark blue arrow represents the GV strongly predicts the exposure, and β_1 the regression coefficient of the GV-exposure association. $\rightarrow$ The light blue arrow represents the β_2 regression coefficient of the exposure-outcome association. $\rightarrow$ The dark and light gray represents β_3 and β_4 , the effects of the confounder (U) on the exposure (X) and the outcome (Y), respectively.

outcome (e.g., disease risk, type 2 diabetes, any health condition). Utilizing these genetic instruments, MR seeks to determine causal effects within observational studies, eliminating the necessity for conventional randomized trials. In other words, MR analytical method helps explore potential causal connections in situations where performing randomized trials is not possible due to ethical dilemmas or practical constraints. **Figure 2** demonstrates the relationships between instrumental variables (Z), exposures (X), and outcomes (Y), while circumnavigating the effects of confounding variables (U). These relationships follow certain assumptions. For example, β_5 demonstrates that no relationship exists between Z and U. Further, β_6 shows that there is no direct relationship between Z and Y, except through X.

Leveraging genetic variants akin to a randomized trial, enables researchers to mimic the randomization process of clinical trials, medicine, pharmacology. Public health, econometric and social sciences. These genetic variants, assumed to be unaffected by confounding variables, are used as proxies for modifiable exposures (e.g., biomarkers, behaviors) to assess causal relationships. Consequently, this approach significantly mitigates biases present in observational studies, allowing for more robust causal inference and empowers researchers to successfully investigate causal relationships between exposures and outcomes. While MR has many advantages, it is not without limitations and challenges. These will be outlined in the section titled “V. Challenges and Limitations of Mendelian Randomization” within this book chapter. A glossary of terms used throughout this chapter is provided in **Box 1**.

Biological pathways: Molecular mechanisms through which genetic variants may influence disease risk.

Causal inference: The process of determining whether a relationship between two variables is causal.

Confounding: Factors that are associated with both the exposure and the outcome, which may distort the observed association.

Exposure: The variable of interest that is hypothesized to have a causal effect on the outcome.

False discovery rate (FDR): The proportion of significant findings that are actually false positives.

Genetic risk score (GRS): A measure that combines the effects of multiple genetic variants to predict an individual's genetic risk for a particular outcome.

Genetic variant: Any specific difference in the DNA sequence between individuals.

Heterogeneity: Differences in the magnitude or direction of effects across different subgroups or studies.

Inherited genetic variation: Genetic variants that are passed down from parents to offspring and are present at birth.

Instrumental variable (IV): A variable used in MR that is associated with the exposure of interest but is not directly associated with confounding factors.

Linkage disequilibrium (LD): Non-random association of alleles at different loci, which can affect the interpretation of genetic associations.

Mediation analysis: Assessing the extent to which the effect of an exposure on an outcome is mediated through an intermediate variable.

Mendelian randomization (MR): A method that uses genetic variants as instrumental variables to assess causal relationships between an exposure and an outcome.

Multiple testing correction: Statistical techniques used to adjust for the increased risk of false positives when multiple hypotheses are tested simultaneously.

Outcome: The variable that is being studied to see if it is affected by the exposure.

Pleiotropy: The situation where a single genetic variant influences multiple phenotypic traits.

Power: The probability of detecting a true effect in a statistical analysis.

Relevance to exposure: Select genetic variants that are strongly associated with the exposure of interest. This ensures that the instrumental variables adequately capture the variability in the exposure and increase the power of the MR analysis to detect causal effects.

Reverse causality: A situation that occurs when the outcome influences the exposure, rather than the exposure influencing the outcome.

Single nucleotide polymorphism (SNP): A common type of genetic variation, occurring when a single nucleotide base is substituted with a different base.

Winner's curse: Genetic associations tend to be overestimated when first discovered; occurs when the strongest associations are selected on the basis of a p-value threshold and lead to over-estimation of the causal association.

Box 1.
Glossary.

2. Source of biases in causal inferences and assumptions for valid Mendelian randomization

Determining the causative risk factors for prevalent conditions is crucial for the development of interventions targeted at treating or managing diseases in clinical studies. The ability to make causal inferences using traditional epidemiological methods is constrained, mainly due to two common types of biases: confounding and reverse causation [9]. Confounding bias occurs when the apparent association between an exposure and an outcome is distorted by the influence of a third variable (confounder) that is connected to both the exposure and the outcome. In simpler terms, bias due to confounding occurs when an external factor falsely impacts the connection between the investigated exposure and the observed outcome. For example: what is the causal effect of coffee consumption on cardiovascular health? Excessive coffee consumption may directly affect cardiovascular health by increasing heart rate and hypertension. However, this observation may be confounded by coffee's indirect effect on a third factor, such as sleep quality, which itself independently influences cardiovascular health. Thus, coffee consumption may negatively affect cardiovascular health either through its direct effect, or by its indirect effect on a third variable, sleep quality. These two effects are unable to be isolated, and thus researchers may find it difficult to determine which exposure is responsible for the outcome. To address confounding bias, various strategies are employed: randomization in experimental studies, matching in observational studies, stratification, and

statistical control. Randomization and matching techniques control for bias by distributing potential confounders evenly across exposure groups; thus, reduce their impact. Stratified analyses, with multivariable regression models and sensitivity analyses, provide additional tools for assessing the influence of confounding variables.

Reverse causation arises when there is a reversal in the perceived cause-and-effect relationship between variables; Y causes X , rather than X causing Y . For example, reduced cardiovascular health may reduce sleep quality, which may in turn lead to increased coffee consumption: in this case, poor cardiovascular health increases coffee consumption, rather than coffee consumption reducing cardiovascular health. To mitigate reverse causation, longitudinal study designs and prospective cohort studies are essential to ensure a clear temporal sequence, minimizing the risk of misinterpreting the cause-and-effect relationship.

Confounding bias and reverse causation may also occur with MR studies. This can lead to misinterpretation of results and inaccurate conclusions about the direction of causation. Consider, for example, an MR study designed to investigate if a genetic variant associated with higher body mass index (BMI) increases the risk of developing type 2 diabetes. This investigation may be confounded by a third factor that also increases BMI, such as a thyroid condition; in this case, the genetic variant may cause high BMI either directly, or indirectly by influencing the thyroid condition. Moreover, reverse causation may occur in the case of individuals whose high BMI results from the early metabolic effects of undiagnosed diabetes; in this case, diabetes would cause high BMI, rather than high BMI causing diabetes. The strategic selection of study designs, coupled with careful consideration of potential confounders, can be used to address reverse causation and confounding bias with multivariable analysis (multivariable MR analysis is beyond the scope of this chapter), thereby strengthening the overall robustness of MR research.

Among the tools used by epidemiologists to address confounding and reverse causation, MR is one of the most robust: it disentangles causal relationships from observational data, utilizing a systematic approach to address the complexities of both reverse causation and confounding bias [10]. First, because genetic variants act as proxies for the exposure of interest, they offer a quasi-randomized setting that minimizes confounding bias. By adjusting levels of the exposure on the outcome, reverse causation and confounding bias are avoided, making it possible to explore the direct impact of genetic variation on the outcome. Second, to address reverse causation, MR effectively establishes temporality and provides a reliable basis for causal inference, as genetic inheritance is fixed at conception.

Genetic variants used in MR studies must satisfy three assumptions related to instrumental variables: *relevance*, *exogeneity* and *independence*. First, the genetic variant should be associated with the exposure of interest (risk factor) to ensure that it has a significant impact on the exposure; this assumption is known as *relevance*. Second, the genetic variant should be unrelated to any confounding factors that might influence the relationship between the exposure and the outcome. This assumption helps isolate the causal effect and is known as *exogeneity*. Third, the genetic variant should be randomly assigned during conception and should not be associated with any factors that might confound the relationship between the exposure and the outcome variable. This is known as the assumption of *independence*. These three assumptions are crucial for the validity of MR studies, as they aim to mimic a RCT using naturally occurring genetic variation to infer causal relationships between an exposure and an outcome. Violation of these assumptions can lead to biased estimates of causal effects. **Figure 3** demonstrates the three assumptions regarding instrumental variables in MR.

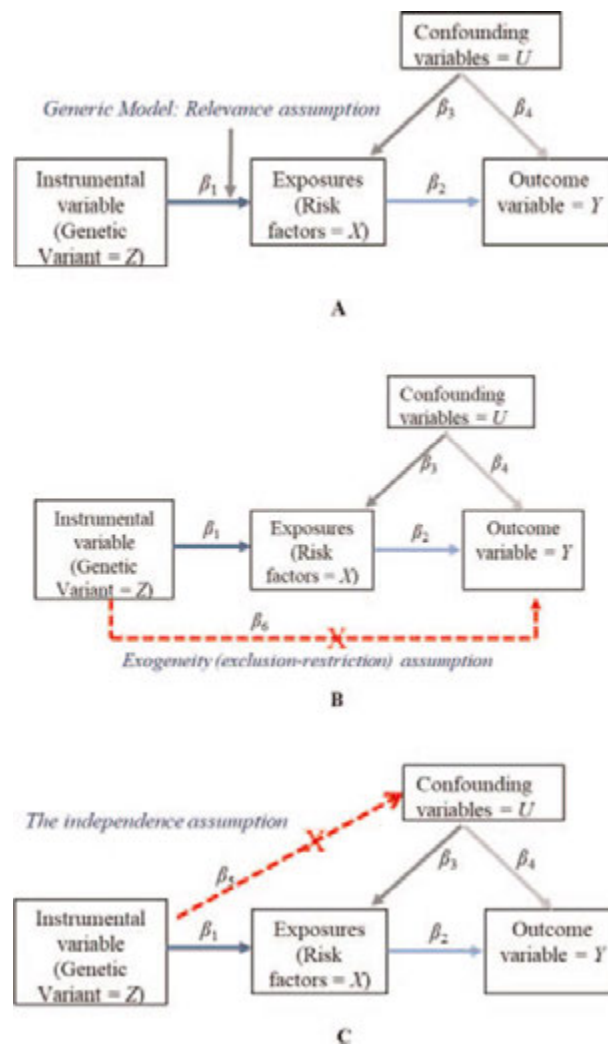


Figure 3.

A–C. Core assumptions in Mendelian Randomization. (A) Relevance assumption: The genetic variant used as an instrument should be associated with the exposure of interest; (B) Exogeneity (exclusion-restriction) assumption: The genetic variant should not be associated with any factors that might confound the relationship between the exposure and the outcome variable, and (C) Independence Assumption: The genetic variant should be unrelated to any confounding factors. --X-- The red dashed arrows indicate pathways that should not exist if the genetic variant (GV) is valid. The line for β_5 means independent of any unobserved confounding variables, also referred to as the independence assumption ($Z \perp U$). The line for β_6 means GV and outcome Y are independent, conditional on exposure X and unmeasured confounders U ($Z \perp Y|X$), also referred to as the exogeneity assumption. $\longrightarrow$ The dark blue arrow represents the GV strongly predicts the exposure, and β_1 the regression coefficient of the GV-exposure association. $\longrightarrow$ The light blue arrow represents the β_2 regression coefficient of the exposure-outcome association. $\longrightarrow$ The dark and light gray represents β_3 and β_4 , the effects of the confounder (U) on the exposure (X) and the outcome (Y), respectively.

The first assumption, *relevance*, can be expressed as follows (**Figure 3A**): let us denote Z as the instrumental variable (genetic variant), X as the exposure of interest (risk factor), and Y as the outcome variable. Relevance implies that there must be a strong association between the genetic variant (Z) and the exposure (X). Statistically, this can be expressed as $\text{Cov}(Z, X) \neq 0$: that is, the covariance between the genetic variant and the exposure should be non-zero, indicating a significant association.

This ensures that the genetic variant is a relevant predictor of the exposure. In statistical terms, this assumption establishes the genetic variant as a valid instrument that has a substantial impact on the exposure. By doing so, researchers can then use the genetic variant as a naturally randomized tool to investigate causal relationships between the exposure and the outcome.

The second assumption, *independence*, implies that the instrumental variable (genetic variant) (Z) is independent of any unobserved confounding variables (U); that is, $Z \perp U$ (**Figure 3B**). Thus, the genetic variant must be statistically independent of any unobserved factors that might confound the relationship between the exposure (X) and the outcome (Y). In other words, the genetic instrument is not influenced by hidden or unmeasured variables that could affect both the exposure and the outcome.

The third assumption, *exogeneity*, implies that the instrumental variable (genetic variant) (Z) is independent of both unobserved confounding variables (U) and the outcome (Y), rather than the exposure (X) (**Figure 3C**). Statistically, this can be expressed as $Z \perp Y|X = E(Y|X)$: the genetic variant must be statistically independent of the unobserved confounders, given the exposure. Thus, the genetic variant must not be influenced by unobserved factors that might confound the relationship between the exposure and the outcome. Determining the size of the causal impact may not always be a primary concern nor indeed feasible; this is especially true for regression outcomes that stem from studies employing diverse data transformations, which may lead to incongruent effect estimates. Nonetheless, when these three assumptions are met, MR can serve the purpose of assessing the presence of a causal relationship.

3. Steps to conduct Mendelian randomization analysis

A comprehensive MR analysis is composed of six primary steps, which are listed in **Table 1**. Each step should be thoroughly addressed in the resulting publication, enabling readers to evaluate model validity. Special reporting guidelines for MR studies are also available to assist in this process [10–13].

Step 1: Define the causal relationship and data sources

Prior to starting a MR study, it is important to first use a robust, well-evidenced hypothesis to identify the causal association that is to be evaluated [14]. Additionally, MR studies may include more than one source of data. For example, a single-sample MR study design may include individual data, or a two-sample MR study design may include both individual data and summary data, such as genome-wide association studies (GWAS) or meta-analyses. Each design may involve multiple steps in the analysis process, as discussed below.

Step	What should be done
1	Define the presumed causal association to be investigated and select data source(s)
2	Select (at least one) genetic variant to be used as the instrumental variable
3	Evaluate the three core assumptions and discuss their applicability
4	Select method of analysis based on study design and conduct MR
5	Assess the validity and robustness of MR findings with sensitivity analysis
51	Report and discuss findings

Table 1.
Six primary steps for conducting a Mendelian Randomization study.

Step 2: Determine criteria for identifying suitable variants/variables

A critical step in conducting a MR study is to determine which genetic variant(s) to include in the analysis [15]. For example, should variants from the same or multiple regions of the genome be included? If the same region is used, results may be under-powered, with all-or-none results: all genes selected may be related to the outcome, or no genes may be related to the outcome. If multiple regions are used, correlated variants may be included, which may lead to instability of results [16]. Multiple genetic variants from different regions may also be considered, such as SNPs [17]. Alternatively, allele scores that summarize multiple genetic variants could be used as instrumental variables [15]. Regardless, the choice of genetic variants should consider both biological and statistical test results that include a pre-specified p-value cutoff.

To successfully implement an MR study, researchers must set criteria that will identify suitable genetic variants to be used as instrumental variables, *Z*. A valid instrumental variable must provide an unbiased estimation of the causal effect of the exposure on the outcome. To do so, it must satisfy the three fundamental assumptions: *relevance, exogeneity, and independence* [18]. The instrumental variable must also be independent of confounders and exhibit minimal collinearity; this ensures that the instrumental variable will provide a valid basis for causal inferences relating the exposure to the outcome. Basic steps and considerations for identifying suitable variants in a MR are summarized in **Box 2**.

Exclusion restriction: Verify that the selected genetic variants only influence the outcome through the exposure of interest, and not through alternative pathways. This principle ensures that the instrumental variables satisfy the exclusion restriction assumption necessary for valid causal inference.

Functional relevance: Consider the biological plausibility of the selected genetic variants in relation to the exposure and outcome of interest. Variants located in or near genes known to be involved in relevant biological pathways are more likely to yield meaningful insights.

Homogeneity: The exposure should have the same effect on all individuals. Though this is not verifiable, one option is to evaluate if the effect is the same across subgroups.

Independence: Select genetic variants that are independent of potential confounders. Ensure that the selected variants are not associated with factors that could bias the causal estimates, such as lifestyle factors, socioeconomic status, or other health-related behaviors.

Multiple instruments: Utilize multiple independent genetic variants as instrumental variables; this will mitigate the potential for weak instrument bias and increase the precision of causal estimates.

Relevance to exposure: Select genetic variants that are strongly associated with the exposure of interest. This ensures that the instrumental variables adequately capture the variability in the exposure and increase the power of the MR analysis to detect causal effects.

Replication: To ensure robustness of findings, validate the association between the selected genetic variants and the exposure in independent datasets or populations.

Sensitivity analyses: Conduct sensitivity analyses, such as MR-Egger regression or leave-one-out analyses, to assess the robustness of results to potential violations of MR assumptions, including pleiotropy or weak instrument bias.

Strength of association: Preferentially select genetic variants with large effect sizes on the exposure, as these variants provide greater statistical power to detect causal effects.

Box 2.
Criteria for identifying suitable variants in a MR.

Regardless of the study design or its method for choosing genetic variants, a comprehensive analysis should evaluate the robustness of the model. It should include sample size calculations for individual data via power analysis (optional) [19], evaluation of assumptions for each instrumental variable, MR estimator, and sensitivity analyses [13].

Step 3: Evaluate core assumptions

Table 2 illustrates a summary of the methods used to evaluate the three fundamental assumptions. As indicated in the table, formal testing is solely applicable to the “relevance” condition. The remaining two conditions can be disproven via sensitivity analyses, yet cannot be proven true [20]. Relevance of Z is tested in the first-stage regression. As a general guideline, for a single exposure regressor, it is recommended that the F-statistic from a joint test should exceed 10 to determine whether all excluded instruments (variables in genetic variant Z_i that are not in exposure, X_i) significantly deviate from zero. Therefore, when presenting genetic variant (instrumental variable (Z_i)) estimates, it is essential to include the F-test results. For situations involving a single instrument and a single exposure regressor, this entails ensuring that the t-value associated with the instrument exceeds approximately $\sqrt{10}$, which is around 3.2, or that the corresponding p-value falls below 0.0016. Hence, choosing suitable genetic variants (instrumental variables) in MR requires careful consideration to ensure the validity of the analysis.

Step 4: MR analysis: Estimation of causal effects

After confirming that genetic variants meet the three assumptions of an instrumental variable—namely, that they exhibit strong association with the exposure (*relevance*), remain independent of confounding factors (*independence*), and solely influence the outcome through the exposure (*exogeneity*)—a suitable method can be chosen to estimate the parameters in the MR analysis. Various methods are able to determine the size of causal effects, with the choice of method depending on the specific practical context. However, it is important to acknowledge that this chapter only presents an introduction to MR: detailed discussion of all methods for estimating causal effects exceeds the scope of this chapter.

There are two overarching approaches to MR methodology: single-sample Mendelian Randomization (SS-MR) and two-sample Mendelian Randomization (TS-MR). Single-sample Mendelian Randomization (SS-MR) is a powerful method used to assess causal relationships between an exposure (e.g., a risk factor or intervention) and an outcome (e.g., a disease) using genetic variants as instrumental variables. The basic principle of MR relies on genetic variants being randomly allocated at conception, similar to a natural RCT. To estimate the causal effects, selection of an appropriate SS-MR method depends on the research question and assumptions. Common methods include, Inverse Variance Weighting (IVW), Wald ratio estimator, and MR-Egger regression, among others.

Conversely, two-sample MR (TS-MR) utilizes GWAS data to estimate the causal effects. There are two ways to use GWAS information as genetic variants; individual level or summary statistics (when individual level GWAS is not publicly available). GWAS-summary statistics are the regression coefficients of SNPs for target outcome variable. Thus, data accessible to the public is driving the advancement of MR techniques utilizing multiple SNP instruments and leveraging GWAS summary statistics for model fitting. See **Table 3** below for an overview of methods in MR, considering both study design and assumptions, and with detailed explanations of each method.

The most frequently used methods are Inverse Variance Weighting (IVW), MR-Egger Regression, Weighted Median Method, MR-PRESSO (Mendelian Randomization Pleiotropy RESidual Sum and Outlier), and Generalized Summary Data Mendelian Randomization (GSMR). Additionally, Bayesian weighted MR, robust adjusted profile score, MRMix, CAUSE, and MRAID are also employed for TS-MR. These methodologies vary in their approaches to three main considerations: instrument selection from available SNPs, modeling and controlling for horizontal pleiotropy, and statistical inference procedures. Notably, MR-Egger can be applied to both GWAS summary statistics and individual-level data to identify biases in meta-analyses and to

Assumptions	Assessment of assumptions	Methods assess the validity
<i>Relevance:</i> Genetic variant is strongly associated with exposure.	<i>Empirically verifiable:</i> Provide association results in study sample(s) via regression, or report from prior evidence. Present biologic support of assumption.	<i>Empirical verification</i> • F-Statistics • Partial R-squared • OR, RR, or RD <i>Prior evidence</i> • Summary statistics from GWAS meta-analyses <i>Biological support</i> • Provide evidence that the genetic variant resides in a gene known to encode the exposure biomarker.
<i>Independence:</i> Genetic variant must not be associated with factors known to confound exposure-outcome association	<i>Not empirically verifiable;</i> testable only to a limited extent: Provide association results of genetic variant with observed variables known to confound observed exposure-disease association. - Provide discussion why variant is unlikely to associate with confounders on the basis of prior biologic knowledge.	<i>Association with confounding</i> • Assess the association between the genetic variant and known confounding variables. • Conduct regression analyses or other appropriate statistical tests to assess the associations between the genetic variant and potential confounders. <i>Biological Plausibility</i> • Use prior biological knowledge between the genetic variant and potential confounders. <i>Literature review</i> • Review existing literature to identify any evidence
<i>Exogeneity:</i> Genetic variant must not affect outcome other than through the exposure.	<i>Not empirically verifiable:</i> Testable only; discuss why variant is unlikely to affect outcome other than through exposure on the basis of prior biologic knowledge.	<i>MR with multiple instruments</i> • Employ MR analyses using multiple genetic variants as instruments for the exposure. If all instruments consistently show a strong association with the exposure and a consistent effect on the outcome, the effects on the outcome are likely mediated through the exposure rather than through alternative pathways. <i>Mediation analysis</i> • Use mediation estimating the direct and indirect effects of the variant on the outcome through the exposure. • If mediation analysis indicates that the majority of the variant's effect on the outcome is mediated through the exposure, the variant is likely to affect the outcome primarily through the exposure.

Table 2.
Methods for evaluating the three assumptions of MR.

Methods	Definition	Technical steps
Inverse-variance weighted (IVW)	IVW is a statistical method commonly used in MR studies to combine the estimates of causal effects obtained from multiple genetic variants (IVs) into a single summary estimate (Hemani et al., 2018).	<i>Inverse- Variance Weighting:</i> In the IVW method, the precision of each causal effect estimate is taken into account. The precision is inversely related to the variance of the estimate. Variance measures how much the estimates of the causal effect for different genetic variants deviate from the true effect. <i>Weight Calculation:</i> The weight assigned to each causal effect estimate is calculated as the inverse of its variance. That is estimates with smaller variances (i.e., more precise estimates) are given higher weights, while estimates with larger variances (i.e., less precise estimates) are given lower weights. <i>Combining Estimates:</i> Once the weights are calculated, they are used to compute a weighted average of the causal effect estimates across all genetic variants. This weighted average represents the overall causal effect of the exposure on the outcome.
The Wald ratio (WR)	This is the simplest approach for estimating a causal effect. The WR is suitable in scenarios where only one SNP serves as a proxy for the target risk factor. The drawback is encountered as it becomes more challenging to assess the assumptions of MR with just a single SNP.	<i>Coefficient Estimation:</i> The coefficient estimated in the linear regression model represents the effect of the genetic variant on the outcome variable. This coefficient reflects the change in the outcome associated with a one-unit increase in the genetically predicted exposure (instrumental variable). <i>Wald Statistic Calculation:</i> The Wald ratio is calculated as the ratio of the estimated coefficient for the association between the genetic variant and the outcome (effect estimate) to the standard error of this coefficient estimate. That is: $WR = \frac{\text{Estimated Coefficient}}{\text{Standard error of coefficient}}$
Weighted median estimate (WME)	Combines individual causal effect estimates from multiple genetic variants, weighting them based on their precision, to derive a robust overall estimate of the causal effect while minimizing the impact of outliers or invalid instrumental variables (Sanderson et al., 2023; Howell et al., 2018).	<i>Weighted Median Calculation:</i> In the WME, the median of the distribution of individual causal effect estimates across all IVs is calculated. This median serves as the overall estimate of the causal effect. <i>Weighting:</i> Unlike other methods like the Inverse-Variance Weighted (IVW) method, the WME assigns equal weight to each instrumental variable. However, it incorporates a weighting scheme based on the precision of the causal effect

Methods	Definition	Technical steps
MR-Egger test	The MR-Egger test is a statistical method used in MR studies to assess the presence of horizontal pleiotropy, which occurs when a genetic variant influences multiple traits independently of the exposure of interest. This approach adjusts the IVW analysis by allowing for a non-zero intercept. A substantial number of instrumental SNPs is necessary; otherwise, the method lacks sufficient power (Sanderson et al., 2023; Howells et al., 2018).	estimates obtained from each genetic variant. <i>Weighting by Precision:</i> The WME assigns higher weights to the causal effect estimates derived from IVs with smaller standard errors, indicating greater precision. Conversely, estimates with larger standard errors are assigned lower weights. <i>Median-based Approach:</i> Provides robustness against potential outliers or invalid IVs. By focusing on the median rather than the mean, extreme or biased estimates from a small subset of IVs have less influence on the overall estimate. <i>Perform MR-Egger Regression:</i> Conduct MR-Egger regression to test for directional pleiotropy, a form of horizontal pleiotropy. The MR-Egger regression allows for the presence of horizontal pleiotropy by relaxing the assumption of the InSIDE (Instrument Strength Independent of Direct Effect) condition. This method estimates a weighted linear regression of the SNP-outcome associations on the SNP-exposure associations, with an intercept term allowing for the presence of pleiotropy. <i>Assess the Intercept Term:</i> Evaluate the intercept term from the MR-Egger regression. A non-zero intercept suggests the presence of directional pleiotropy. This implies that at least one genetic variant affects the outcome directly, violating the instrumental variable assumptions. <i>Interpret the Results:</i> Interpret the results cautiously, considering the assumptions underlying MR analyses and the potential implications of horizontal pleiotropy on the validity of causal inference.
Maximum Likelihood Estimation (MLE)	MLE estimates the causal effect of an exposure on an outcome in MR studies, leveraging genetic instrumental variables to infer causality while accounting for potential sources of bias and uncertainty (Howells, et al., 2018).	<i>Specify the Likelihood Function:</i> Formulate the likelihood function that represents the probability of observing the genetic data given the parameters of the model. <i>Maximize the Likelihood:</i> Use numerical optimization techniques to maximize the likelihood function with respect to the parameters of the model. This involves iteratively adjusting the parameter values to find the values that maximize the likelihood of observing the genetic data.

Methods	Definition	Technical steps
		<i>Estimate Causal Effect:</i> Once the likelihood function is maximized, estimate the causal effect of the exposure on the outcome using the parameter estimates obtained from the optimization process. <i>Assess Uncertainty:</i> Calculate confidence intervals or standard errors for the estimated causal effect to quantify uncertainty around the point estimate. This can be done using methods such as bootstrapping or by estimating the asymptotic covariance matrix of the parameter estimates.
Two-stage Ordinary Least Squares (2S-OLS)	Two-stage Ordinary Least Squares (OLS) regression is a commonly used method in MR studies for estimating the causal effect of an exposure on an outcome using genetic instrumental variables.	<i>First Stage Regression:</i> In the first stage, regress the exposure variable (X) on the genetic instruments (Z) using OLS regression. The regression equation in first stage: $X_i = \beta_0 + \beta_1 Z_i + e_{1i}$ Where β_0 denotes intercept and β_1 is the slope. Z_i stands for the SNP instrument genotype for individual i, and e_{1i} random error term assumed to be independent among individuals and normally distributed with a common variance and mean zero. <i>Instrument Strength Assessment:</i> Assess the strength of the genetic instruments using statistical measures such as the F-statistic or the partial R-squared. <i>Second Stage Regression:</i> In the second stage, regress the outcome variable (Y) on the predicted values of the exposure variable ($\hat{X}_i$) obtained from the first stage regression, along with any relevant covariates. The regression equation for the second stage: $y_i = \beta_{01} + \beta_2 \hat{X}_i + e_{2i}$ while $\hat{X}$ represents the predicted value of i individual of the exposure from the first stage regression, β_{01} and β_2 represent estimated intercept and causal effect, and e_2 is random error term. As assumed in the first stage, error in the second stage errors to be independent among individuals and normally distributed with a common variance and mean zero.

Methods	Definition	Technical steps
MR-PRESSO (Mendelian Randomization Pleiotropy RESidual Sum and Outlier)	This statistical method is used to detect and correct for horizontal pleiotropy, a violation of MR assumptions where genetic variants influence the outcome through pathways other than the exposure of interest.	<i>Outlier Detection:</i> MR-PRESSO uses a residual sum and outlier (RSO) statistic to detect outliers by comparing the observed residuals of each SNP to the expected distribution of residuals. <i>Outlier Correction:</i> Once potential outliers are identified, MR-PRESSO proceeds to correct for their influence on the MR estimates. Outliers can be corrected through a correction step where the causal effect estimates are recalculated after removing the outlier SNPs. <i>Residual Sum and Outlier Test:</i> Conduct the MR-PRESSO test to formally assess the presence of horizontal pleiotropy. This involves regressing the SNP-outcome associations on the SNP-exposure associations and examining the distribution of the residuals. The test assesses whether the residuals exhibit excess variability beyond what is expected under the null hypothesis of no pleiotropy.
Generalized Summary Data Mendelian Randomization (GSMR)	This method is used for parameter estimate in MR to analyze summary-level data from genome-wide association studies (GWAS) for both the exposure and the outcome of interest.	<i>Linkage Disequilibrium (LD) Estimation:</i> Calculate pairwise linkage disequilibrium (LD) between genetic variants (SNPs) using reference LD panels (e.g., 1000 Genomes Project data). LD estimation is necessary for GSMR to account for correlations between SNPs due to LD. <i>GSMR Analysis:</i> Use the summary statistics and LD information to perform GSMR analysis. GSMR typically involves fitting a weighted linear regression model to estimate the causal effect of the exposure on the outcome. The model includes SNP-exposure associations as the independent variable and SNP-outcome associations as the dependent variable, with LD information incorporated into the analysis.

Table 3.
Overview of MR methods.

detect biases stemming from horizontal pleiotropy in two-sample MR studies [21]. Furthermore, for individual-level data, methods such as MVMR, MR-GENIUS, sisVIVE, and Adaptive LASSO are utilized [9].

Step 5: Assess robustness of findings with sensitivity analyses

Following the parameter estimation, the next step is to conduct sensitivity analyses. Sensitivity analyses help researchers assess the validity and robustness of MR findings and provide insights into the potential impact of biases or violations of assumptions. In order to enhance the reliability of causal inference, MR results are commonly presented alongside sensitivity analyses that account for different patterns of pleiotropy. Pleiotropy occurs when one genetic variant is associated with multiple phenotypes and can be either horizontal or vertical. Horizontal pleiotropy is a violation of the exogeneity assumption and occurs when the genetic variant influences the outcome directly; i.e., the variant is *independent* of the exposure. Vertical pleiotropy is a spurious, or false, pleiotropy that occurs when a genetic variant affects other traits which influence the outcome but is *dependent* on the exposure. As such, this genetic variant can be used to estimate the causal influence between traits [22]. Additional approaches are performed to assess model robustness and detect potential violations of MR assumptions, such as pleiotropy or weak instrument bias.

Sensitivity analyses are performed automatically in many applications, as in MR-Base software [23]. Common sensitivity analyses and diagnostics in MR are heterogeneity tests, leave-one-out (LOO) analyses and funnel plots. Heterogeneity can be calculated for the inverse variance weighted (IVW) and MR-Egger tests, and this can be used to navigate between models of horizontal pleiotropy [24]. To assess whether the MR estimate is influenced or biased by a single SNP with potentially significant horizontal pleiotropic effects, we can recalibrate the effect by removing one SNP at a time in sequence. Identifying SNPs that, when excluded, cause a substantial alteration in the estimate can offer valuable insights into the estimate's sensitivity to outliers [22]. As frequently employed in meta-analysis applications, funnel plots represent an alternative approach for sensitivity analysis. Here, the estimate for a specific SNP is plotted against its precision. Detecting asymmetry in the funnel plot may suggest violations of the exogeneity assumption due to horizontal pleiotropy.

Step 6: Interpretation and reporting

Interpreting causal effects in MR analysis requires careful consideration of the underlying assumptions, methodological choices, and potential sources of bias [25]. Once the parameters have been estimated, the interpretation phase becomes essential for drawing meaningful conclusions from the results. However, potential limitations, such as pleiotropy, may be present and must be acknowledged. The STROBE-MR provides a comprehensive guideline for MR analyses and reporting ([13]; <https://www.strobe-mr.org>).

4. Software and tools for Mendelian randomization analysis

To conduct MR analysis effectively, researchers rely on a range of specialized software and packages tailored to the unique demands of this methodology. The availability of sophisticated software solutions has significantly enhanced the precision, efficiency, and reliability of MR, empowering researchers to unravel complex causal pathways with unprecedented clarity. Here, we provide an overview of some prominent MR software and packages, see **Table 4**. Performing MR analysis requires careful consideration of study design, data sources, and analytical methods.

R/RStudio [26]	Python [27]	Other software
<i>MR-Base</i> : A comprehensive platform that provides access to a wealth of summary-level data for MR analysis. It hosts curated databases of genome-wide association studies (GWAS) and summary statistics, along with tools for performing various MR analyses, including inverse variance-weighted (IVW), MR-Egger, and weighted median methods.	<i>MendelianRandomization</i> : Conducts various MR methods, including inverse-variance weighted (IVW) analysis, weighted median, MR-Egger regression, and more. This package often serves as the go-to tool for many researchers due to its comprehensive set of features.	<i>mrrobust (Stata)</i> : It provides functionalities for several MR methods, including inverse-variance weighted (IVW) estimation and robust alternatives. <i>Ivregress or ivreg2 (Stata)</i> : These procedures fit a regression via 2SLS but also via GMM (generalized method of moment) when 2SLS is used. It can directly show the estimates of both the first and second stage of least squares. <i>simmr (BioRxiv)</i> : This can be useful for researchers developing new MR methods or evaluating the performance of existing ones.
<i>TwoSampleMR</i> : Designed specifically for conducting MR analyses using summary-level data from GWAS. It offers a range of methods for MR analysis, including IVW, MR-Egger, weighted median, and MR pleiotropy residual sum and outlier (MR-PRESSO) tests. This also facilitates sensitivity analyses and provides tools for assessing horizontal pleiotropy.	<i>PyMendel</i> : Offers functionalities for genetic analysis, including MR analysis. It provides tools for data preprocessing, performing MR analysis using various methods, and generating plots to visualize results.	<i>PROC SYSLIN and PROC QLIM (SAS)</i> : These procedures allow for the implementation of instrumental variable (IV) regression methods like two-stage least squares (2SLS) – a technique relevant to MR analysis. However, customization might be needed for specific MR methods. <i>PROC REG (SAS)</i> : This procedures in SAS can be used to estimate the causal effect (outcome) of an exposure (mediator) on an outcome variable, conditioned on a genetic instrument (exposure). However, it would not account for specific MR assumptions
<i>MendelianRandomization</i> : This offers functionalities for performing both single and multivariable MR analyses. It includes methods for estimating causal effects using summary-level data and provides options for sensitivity analyses and model diagnostics.	<i>Genal</i> : User-friendliness and streamlining complex analyses. This provides us help for data cleaning for single nucleotide polymorphism (SNP) data, association testing, polygenic risk scoring. MR analyses with various methods. Offers core functionalities from established R packages like TwoSampleMR and MR-PRESSO.	<i>MendelianRandomization.jl (Julia)</i> : This is a Julia package specifically designed for MR analysis. It provides functionalities similar to those available in Python and R packages for MR analysis.
<i>MR-Robust</i> : This is particularly in the presence of violations of MR assumptions such as horizontal pleiotropy. It implements robust methods, including robust adjusted profile score (RAPS) and MR-Egger robust regression, to provide more reliable causal estimates.	<i>MeRP</i> : Streamlines the curation of genetic variants from GWAS catalog. There are three components of the MeRP pipeline: i) Pulling genetic variants from public data, ii) filtering to satisfy MR conditions, and iii) the causal effect-score calculation for MR analysis.	

Table 4.
Software for MR analyses.

Here, we outline a general guideline on using specific software, such as MR-Base and TwoSampleMR, for conducting MR analysis:

4.1 Data preparation

Obtain summary-level data for the exposure, outcome, and genetic instruments from reputable sources or conduct GWAS to generate summary statistics. Ensure data quality and consistency across datasets, including harmonization of variants and allele frequencies.

4.2 Instrument selection

Identify genetic instruments (i.e., instrumental variables) associated with the exposure of interest and independent of confounders using methods like genome-wide significance thresholding and LD clumping. Check for pleiotropy and assess instrument strength using methods like F-statistic and MR-Egger intercept.

4.3 MR analysis

Use MR-Base platform or TwoSampleMR package to perform MR analysis, selecting appropriate methods based on the research question and assumptions. Conduct primary analysis using methods like IVW or weighted median to estimate causal effects. Perform sensitivity analyses, including MR-Egger regression and leave-one-out analysis, to assess the robustness of results.

When all three MR assumptions are upheld, the interpretation of results becomes meaningful, enabling the evaluation of regression coefficients and the testing of associations between genetic variants, exposure data, and outcome data. These interpretations also involve assessing the magnitude and direction of causal effects estimated by the chosen method. Any violation of the three assumptions can undermine the validity of causal inference. In practical application, a thorough assessment of the assumptions becomes imperative. Given that exogeneity and independence assumptions pertain to the absence of specific types of confounding, their verification remains unattainable, requiring practitioners to exercise caution in the planning of MR analyses.

For single-sample MR, where genetic variants serve as instrumental variables, the estimated causal effects reflect the impact of the exposure on the outcome. However, interpreting these effects requires caution, as they may be biased. Sensitivity analyses, such as leave-one-out tests, can help assess the robustness of results to potential violations of the core assumptions. In two-sample MR, where GWAS data are leveraged, interpreting causal effects involves considering the strengths and limitations of using summary statistics from separate samples. While this approach can enhance statistical power and generalizability, it also introduces challenges such as horizontal pleiotropy, where genetic variants influence the outcome through pathways other than the exposure of interest. Methods like MR-Egger regression can help detect and adjust for such biases, but their interpretation also requires careful consideration of underlying assumptions.

Furthermore, interpreting MR results entails considering the broader context of the research question and existing evidence. Causal effects estimated through MR should be interpreted in conjunction with findings from observational studies, RCT, and other types of evidence. Consistency across different study designs strengthens

causal inference and increases confidence in the results. Thus, researchers should provide detailed descriptions of methods used, including software and packages, the MR estimates with the magnitude and direction of causal effects, along with results of sensitivity analyses. Implications of findings and potential avenues for future research should also be discussed. By following these steps and leveraging specialized MR software and packages, researchers can conduct rigorous MR analyses to investigate causal relationships between exposures and outcomes, contributing to advancements in epidemiology and genetics.

Overall, interpreting causal effects in MR analysis requires a nuanced understanding of the methods used, the assumptions underlying them, and the broader context of the research question. By carefully considering these factors, researchers can draw reliable conclusions about the causal relationships between exposures and outcomes in observational data. For thorough reporting of MR studies, consider using EQUATOR network (<https://www.equator-network.org/reporting-guidelines/>) and checklists, such as STROBE-MR for reporting MR studies [13, 28].

5. Challenges and limitations of Mendelian randomization

MR faces several challenges and limitations that warrant consideration. First, MR is constrained by its reliance on large sample sizes. Often, the genetic variants that serve as proxies for exposures or traits account for only a minute fraction of the overall variability in those exposures or traits. Consequently, to derive precise risk estimates, researchers typically require datasets comprising thousands of samples [29].

Second, MR analysis can be complicated by linkage disequilibrium, wherein multiple genetic variants are inherited jointly - these variants may be heterogeneous in their impact on the exposure or outcome, which can potentially undermine the validity of the findings. To ensure the robustness and reliability of MR results, this factor must be meticulously considered and adjusted for.

A third limitation is population stratification, which may occur in studies involving mixed or stratified populations. In this case, spurious associations may emerge due to the correlation between the genetic variant and the outcome with the ancestral background. To mitigate this issue, methodologies have been developed to either confine analyses to homogeneous populations, or to adequately control for population stratification using techniques such as principal components analysis or linear mixed models [30].

Fourth, especially present in single-sample MR studies, is the “winner’s curse” phenomenon: when a genetic association between an exposure and an outcome is first discovered, the significance of that association tends to be overestimated compared to later studies. When a sample subject to the “winner’s curse” is utilized for both the discovery and the MR analysis for constructing the genetic instrument, there is a risk of biasing the instrumental variable and exposure estimates upwards. This limitation may be circumvented by employing a two-sample MR approach.

A fifth limitation stems from trait heterogeneity, horizontal pleiotropy, and linkage disequilibrium. These complexities undermine the underlying assumptions of MR analyses, and necessitate a deep understanding of genetic variants and biological mechanisms. To address these issues, specialized techniques, such as MR-Egger regression and weighted median approaches, have been developed to detect and account for pleiotropy [29].

Sixth, because MR relies on assumptions about the biological pathways that link genetic variants to exposures and outcomes, unanticipated complexity in the underlying biology may lead to misinterpretation of causal relationships. Integrating biological knowledge and conducting robust sensitivity analyses can help to mitigate this limitation.

Despite these constraints, MR studies remain capable of offering valuable insights into causal connections between exposures and outcomes, particularly in scenarios where alternative study designs, such as randomized controlled trials, are impractical or unethical. It is essential for researchers to meticulously weigh the assumptions and constraints of MR and supplement their findings with additional evidence to fortify causal inference. This approach ensures a comprehensive understanding of the relationships under investigation. Several researchers have provided more details on the limitations of MR; [31–34] the utility of MR is contingent upon navigating these limitations through robust methodologies and careful consideration of genetic and population factors.

6. Future directions in applications of Mendelian randomization

The future MR holds significant promise as advancements in genomics and data science continue to evolve. One key direction is the integration of multi-omics data, including epigenomics, transcriptomics, and metabolomics, to provide a more comprehensive understanding of the causal pathways linking genetic variants to complex traits and diseases [30, 33]. This holistic approach can uncover novel biomarkers and therapeutic targets, enhancing precision medicine.

Another promising avenue is the application of MR in diverse populations. Traditionally, MR studies have been predominantly conducted in European populations, limiting the generalizability of findings [20, 21]. Expanding MR research to include diverse ethnic groups will improve the robustness of causal inferences and ensure the benefits of genomic discoveries are equitably distributed.

The development of sophisticated statistical methods and machine learning algorithms also presents a future direction for MR. These advancements can improve the detection and correction of pleiotropy, a major challenge in MR, by distinguishing true causal effects from confounding influences. Enhanced computational tools will facilitate more accurate and reliable MR analyses, making them accessible to a broader range of researchers [11, 31].

Finally, the integration of MR with electronic health records (EHRs) and large-scale biobanks, such as the UK Biobank [20], and All of Us Research Program, will provide unprecedented opportunities for causal inference in real-world settings. This will enable the investigation of gene-environment interactions and the causal effects of lifestyle and environmental factors on health outcomes. Furthermore, the ethical and regulatory aspects of MR will gain increasing attention. As MR becomes more influential in shaping public health policies and clinical guidelines, it is crucial to address ethical considerations related to genetic privacy, data sharing, and the responsible communication of findings to the public.

7. Conclusion

Mendelian Randomization represents a powerful tool for causal inference, providing valuable insights into the relationships between genetic variants, exposures, and

health outcomes. As the field advances, addressing biases, enhancing methodological approaches, and incorporating diverse data sources will be key to unlocking its full potential. By fostering inclusivity and ethical integrity, MR will continue to contribute significantly to the understanding of disease etiology and the improvement of public health interventions. Continuing advancement in the fields of genomics and data science will ensure that MR remains an indispensable tool for causal analysis.

Acknowledgements

We would like to thank Brian Egan, PhD for his thoughtful suggestions and edits of the chapter contents. His contributions were much appreciated.

Conflict of interest

The authors declare no conflict of interest.

Author details

Hayrettin Okut^{1,2*} and Rosey Zackula^{1,3}

1 Office of Research, University of Kansas School of Medicine, Wichita, KS, United States

2 Department of Population Health, University of Kansas School of Medicine, Wichita, KS, United States

3 Department of Psychiatry and Behavioral Science, University of Kansas School of Medicine, Wichita, KS, United States

*Address all correspondence to: hokut@kumc.edu

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Harrison S, Howe L, AR D. In: Division RAE, editor. Making Sense of Mendelian Randomisation and its Use in Health Research: A Short Overview. Cardiff Wales: Public Health Wales NHS Trust; 2020
- [2] Evans DM, Davey Smith G. Mendelian randomization: New applications in the coming age of hypothesis-free causality. Annual Review of Genomics and Human Genetics. 2015;**16**:327-350
- [3] Plotnikov D, Guggenheim JA. Mendelian randomisation and the goal of inferring causation from observational studies in the vision sciences. Ophthalmic & Physiological Optics. 2019;**39**(1):11-25
- [4] Zuber V et al. Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. American Journal of Human Genetics. 2022;**109**(5):767-782
- [5] Gagliano Taliun SA, Evans DM. Ten simple rules for conducting a mendelian randomization study. PLoS Computational Biology. 2021;**17**(8): e1009238
- [6] Bandres-Ciga S, Noyce AJ, Traynor BJ. Mendelian randomization-a journey from obscurity to center stage with a few potholes along the way. JAMA Neurology. 2020;**77**(1):7-8
- [7] Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet. 1986;**1**(8479):507-508
- [8] Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. Bone Marrow Transplantation. 1991;**7** (Suppl 3):9-12
- [9] Hartley AE et al. A guide for understanding and designing Mendelian randomization studies in the musculoskeletal field. JBMR Plus. 2022; **6**(10):e10675
- [10] Boef AG, Dekkers OM, le Cessie S. Mendelian randomization studies: A review of the approaches used and the quality of reporting. International Journal of Epidemiology. 2015;**44**(2): 496-511
- [11] Burgess S et al. Guidelines for performing Mendelian randomization investigations: Update for summer 2023. Wellcome Open Research. 2019;**4**:186
- [12] Sekula P et al. Mendelian randomization as an approach to assess causality using observational data. Journal of the American Society of Nephrology. 2016;**27**(11):3253-3265
- [13] Skrivankova VW et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): Explanation and elaboration. BMJ. 2021;**375**:n2233
- [14] Adam D. The causation detector. Nature. 2019;**576**:196-199
- [15] Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. International Journal of Epidemiology. 2013;**42**(4): 1134-1144
- [16] Karageorgiou V et al. Sparse dimensionality reduction approaches in Mendelian randomisation with highly correlated exposures. eLife. 2023;**12**: e80063
- [17] Davies NM, Holmes MV, Davey Smith G. Reading Mendelian

randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ*. 2018; **362**:k601

[18] de Leeuw C et al. Understanding the assumptions underlying Mendelian randomization. *European Journal of Human Genetics*. 2022;**30**(6):653-660

[19] Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *International Journal of Epidemiology*. 2013;**42**(5):1497-1501

[20] Sanderson E et al. Mendelian randomization. *Nature Reviews Methods Primers*. 2022;**2**(6)

[21] Boehm FJ, Zhou X. Statistical methods for Mendelian randomization in genome-wide association studies: A review. *Computational and Structural Biotechnology Journal*. 2022;**20**: 2338-2351

[22] Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Human Molecular Genetics*. 2018;**27**(R2):R195-R208

[23] Hemani G et al. The MR-base platform supports systematic causal inference across the human phenome. *eLife*. 2018;**7**:e34408

[24] Bowden J et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. *International Journal of Epidemiology*. 2018;**47**(4):1264-1278

[25] Haycock PC et al. Best (but oft-forgotten) practices: The design, analysis, and interpretation of Mendelian randomization studies. *The American Journal of Clinical Nutrition*. 2016; **66** (4):965-978

[26] Yavorska OO, Burgess S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *International Journal of Epidemiology*. Dec 1 2017;**46**(6):1734-1739

[27] Kinsley H, Kukiela D. Neural Networks From Scratch in Python. Sentdex: Kinsley Enterprises Inc; 2020

[28] Au Yeung SL, Gill D. Standardizing the reporting of Mendelian randomization studies. *BMC Medicine*. 2023;**21**(1):187

[29] Go T-H, Kang DR. Basic concepts of a Mendelian randomization approach. *Cardiovascular Prevention and Pharmacotherapy*. 2020;**2**(1):24-30

[30] Zheng J et al. Recent developments in Mendelian randomization studies. *Current Epidemiology Reports*. 2017; **4**(4):330-345

[31] Davey Smith G, Hemani G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*. 2014;**23**(R1): R89-R98

[32] Davey Smith G, Paternoster L, Relton C. When will Mendelian randomization become relevant for clinical practice and public health? *JAMA*. 2017;**317**(6):589-591

[33] Howell AE et al. Use of Mendelian randomization for identifying risk factors for brain tumors. *Frontiers in Genetics*. 2018;**9**:525

[34] Vander Weele TJ et al. Methodological challenges in mendelian randomization. *Epidemiology*. 2014; **25**(3):427-435

Environmental Pollutant-Based Polygenic Risk Score: A Genetic Surrogate for Environmental Pollutants Assessment

Mulong Du, Junyi Xin, Silu Chen and Meilin Wang

Abstract

Accurate measurement of environmental pollutants is necessary for precision health risk assessment. However, current methods for detecting and profiling humans' internal exposure to pollutants remain prohibitively costly, time-consuming, and energy-intensive. Polygenic risk scores (PRSs), which aggregate effects of genetic variants across the genome, have been widely applied in predicting phenotypes and could serve as genetic surrogates for traits of interest. This includes a genetic proxy relevant to environmental exposure to pollutants, facilitating disease prediction and risk evaluation. This chapter introduces a state-of-the-art technology: the environmental pollutant-based PRS (EpPRS), which harnesses the complete genetic and exposomic data. Furthermore, we have developed an adverse outcome pathway (AOP) framework aimed at assessing the effects of contaminants on the risk of candidate diseases. In addition, EpPRS, as a surrogate of genetically predicted exposure, can be effectively utilized in various analyses, including causal inference, mediation analysis, and phenome-wide association studies. In conclusion, EpPRS can act as a robust proxy to assess internal pollutant exposure. The application of the EpPRS in disease risk assessment elucidates the mechanisms linking exposure and disease, thereby providing a foundation for developing strategies to control environmental pollutants.

Keywords: environmental contaminant, polygenic risk score, genetic surrogate, risk assessment, environmental health

1. Introduction

Environmental contaminants are an ever-present concern in modern human environments, characterized by their diversity and widespread nature. The development of sophisticated detection techniques has heightened awareness of the risks to both environmental and human health posed by these chemicals. Despite advancements, accurately assessing health risks remains a formidable challenge on account of the inherent persistence, high toxicity, long-distance mobility, and bioaccumulation

of pollutants. Although some countries have implemented bans on various environmental contaminants, their production and use continue in others [1–3]. Additionally, recent climate change has facilitated the unrestricted release of pollutants from ice and soil sediments [4]. Persistent organic pollutants (POPs), for instance, are synthetic chemicals known for their environmental persistence, bioaccumulate in the food chain, and significant health hazards [5]. The toxicity of POPs is particularly magnified for apex predator within the food chain. POPs remain omnipresent in diverse food items, like milk, meat, fish, and dairy products, inflicting chronic damage to humans due to lifelong exposure [6].

Traditional risk assessment models have proven inadequate for evaluating the human health hazards posed by the multitude of chemical compounds and environmental contaminants. To address these challenge, traditional risk assessment methods are undergoing significant revisions. Recently, the adverse outcome pathway (AOP) has been proposed as a novel approach to incorporate correlations between adverse outcomes and risk assessments across different biological organizational levels [7]. The AOP framework commences with a molecular initiating event (MIE), where a compound interacts with a biological target, triggering a series of higher-level events (such as subcellular, cellular, system, or individual level) that culminate in adverse health outcomes. However, for environmental pollutants, accurate exposure assessment remains crucial. The detection of human internal exposure to pollutants remains costly, time-consuming, and resource-demanding. Additionally, limitations in resources for monitoring and management necessitate the development of cost-efficient alternative approaches for evaluating environmental exposure levels.

The polygenic risk score (PRS) has emerged as an important tool in the genetic research of complex trait diseases, complementing genome-wide association studies (GWAS) [8]. PRS aggregates the effects of multiple genetic variants across the genome to predict the genetic predisposition of individuals to certain traits or diseases. Genetic variations in environmental response genes can significantly influence the status of pollutant biomarkers and/or the relationships between pollutant exposure and biomarkers [9, 10]. This implies that individual genotypes might influence the impacts of environmental pollutant exposure. Some investigations have integrated environmental genomic data to explore genetic susceptibility to diseases influenced by the environment, identify genes vulnerable to chemical damage, and pinpoint functionally significant single-nucleotide polymorphisms (SNPs) in genes responding to environmental stimuli [11]. Taking inspiration from the PRS approach, we suggest leveraging genetic tools from environmental genomics to predict individual risks due to contaminant exposure.

Given the extensive research conducted on environmental exposures, the genotoxicity of POPs on humans warrants attention. Polychlorinated biphenyls (PCBs) and organochlorine pesticides (OCPs) were among the first substances enlisted under the Stockholm Convention in 2004 [12], following polybrominated diphenyl ethers (PBDEs) added in 2009 (<https://www.pops.int/>). Despite long-standing prohibitions on POPs, certain countries continue to produce and utilize these substances, and climate change has further contributed to their unrestricted release from ice and soil sediments [4]. Utilizing POPs as a case example, we developed an environmental pollutant-based polygenic risk score (EpPRS) by means of genetic methods. This method aims to enhance understanding of biological variation across and within individuals, offering another option for evaluating environmental pollutant exposure when direct measurements are unavailable. In order to thoroughly illustrate the

toxic pathways connecting POP exposure and disease risk, we suggest employing the EpPRS as a genetic substitute to evaluate the health risks posed by environmental contaminants within the AOP framework.

The integration of comprehensive genetic and exposomic data through the EpPRS offers a novel approach to improve the accuracy of health risk assessments related to environmental pollutants. This innovative method combining genetic tools and environmental genomics lays the groundwork for more effective strategies in pollutant control and public health protection. By providing detailed insights into the biological mechanisms and pathways affected by environmental exposures, the EpPRS framework has the potential to significantly advance our understanding of how pollutants contribute to disease risk and progression. Moreover, this approach can facilitate the development of personalized intervention strategies and inform policy decisions aimed at mitigating the adverse health effects of environmental pollutants.

In conclusion, the challenge of accurately assessing the health risks posed by environmental contaminants remains significant. However, through the development and application of novel methodologies such as the EpPRS and the AOP framework, we can enhance our ability to predict and manage these risks. The continued integration of genetic and exposomic data will be crucial in advancing our understanding of the complex interactions between environmental exposures and human health, ultimately leading to more effective public health strategies and interventions.

2. Materials

2.1 Research participants

Individuals were enlisted from a study based on community. From two separate populations, 714 individuals were selected, each having both POP exposure and GWAS data. The training phase consisted of 410 participants and the validation stage comprised 304 subjects. Subjects who have serious cognitive impairment or psychiatric disorders were excluded from this study. The type 2 diabetes mellitus (T2DM) history was documented through self-reported diagnosis, in which the participant responded “yes” for either “Have you ever used the hypoglycemic drug?” or “Fasting blood glucose ≥ 7.0 mmol/L?” [13].

The age distribution of participants in both datasets was similar (mean = 60.58 and 59.78). There was a higher prevalence of females in the validation stage (43.4 vs. 64.5%). Diabetes prevalence stood at 19.1%, marginally surpassing prior studies [14]. All research subjects signed informed consent, and the study was approved by the Nanjing Medical University Institutional Review Board.

2.2 Determination of persistent organic pollutants

The blood plasma samples obtained from the participants subjected to analysis for the presence of persistent organic pollutants (POPs), comprising 2 OCPs, 15 PCBs, and 5 PBDEs, using gas chromatography-tandem mass spectrometry (GC-MS/MS). Levels under half the detection limit (LOD) were treated as under LOD. POP levels were transformed into standard normal distributions via inverse normalization. The concentrations of the 2 OCPs, 15 PCBs, and 5 PBDEs were combined to evaluate their overall effects.

2.3 Genome-wide association study and heritability estimates for POPs

Using Illumina Human Omni ZhongHua Bead Chips, genotyping was performed in accordance with the protocols detailed in our previous studies [15]. The SNPs and samples underwent consistent quality control protocol. Exclusions were on the basis of discrepancies in sex, genetic relatedness ($PI_HAT > 0.25$), or unanticipated duplications. Non-genotyped SNPs were imputed using IMPUTE2 according to the 1000 Genomes Project. SNPs were excluded if they met any of the following criteria: (i) call rate $< 95\%$; (ii) minor allele frequency (MAF) < 0.05 ; (iii) localized on the sex chromosome; (iv) imputation info score < 0.3 ; or (v) Hardy-Weinberg equilibrium P value $< 1 \times 10^{-6}$. A total of 3,938,474 autosomal SNPs were kept for statistical analysis. Principal component analysis (PCA) was employed to evaluate stratification of populations and outlier points, referencing the 1000 Genomes Project (March 2012).

Using PLINK 1.90, the correlation between POPs and SNPs was analyzed with an additive model that accounted for eigenvectors, age, and sex. The quantile-quantile and Manhattan plots, which were on the basis of $\log_{10}$ -transformed P values, were generated with the *CMplot* R package. No evidence was found to suggest that P values had been systematically inflated, indicating that no significant residual population structure was present.

The genomic-relatedness-matrix restricted maximum likelihood (GREML) model, implemented in genome-wide complex trait analysis software (GCTA v1.94.1), was conducted on individual-level GWAS data to estimate SNP-based heritability (h^2_{SNP}) for any POP [16]. Genetic relationships between individual pairs were determined using all autosomal markers with $MAF \geq 0.05$. Each POP dataset was analyzed later with a linear model, and GREML was employed to estimate the variance explained by genome-wide SNPs, identical to the h^2_{SNP} . Covariates included age, sex, and the top three principal components (PCs). The first three principal components could explain approximately 80% of the total variance. Due to the small sample size, which could reduce statistical efficiency, GWAS sets with estimates below 0.01 were excluded. The heritability estimates ranged from 0.036 for BDE99 to 0.503 for BDE100.

3. Construction of EpPRS

The PRS was calculated using the formula: $PRS = \sum_{i=1}^n \beta_i SNP_i$, where SNP_i is the number of risk alleles (0, 1, or 2) of each SNP, and β_i is the coefficient of linear regression derivative of the POP GWAS in the training phase. Data of individual-level genotypes from the validation phase were used to construct EpPRS (**Figure 1**). To identify the most effective, we created the PRS of each POP using three diverse techniques: Lasso, PRS-CS, linkage disequilibrium (LD) clumping, and P value thresholding (C + T). In total, we created 14 EpPRSs were generated for each POP exposure. All EpPRSs were statistically positive and were found to be correlated with their corresponding concentrations observed during the training phase and were thus selected for deeper analysis. The specific methods are described below.

3.1 LD clumping and P value thresholding (C + T)

We employed LD clumping to eliminate all non-independent SNPs, utilizing the 1000 Genomes Project East Asian populations as a reference. According to the

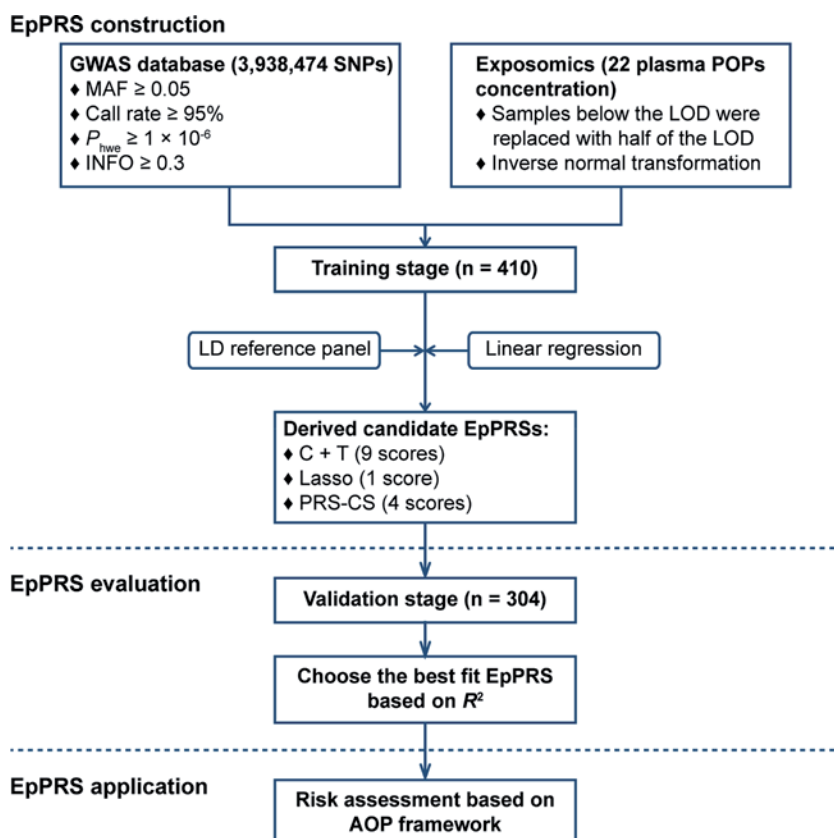


Figure 1.
Flowchart depicting the construction, evaluation, and application of the EpPRS.

summary statistics of POPs, PLINK software was used to identify three groups of variants [17]. The region size was set to 1 Mb and the LD r^2 to 0.001, 0.01, and 0.1, with three different P value thresholds (1×10^{-5} , 1×10^{-4} , and 1×10^{-3}). The threshold of P value that resulted in the greatest R^2 among the nine scores was selected.

3.2 Lasso

An optional approach to C + T PRS construction utilizing complete summary statistics is “Lasso” [18]. The PRS weights are calculated in this approach utilizing summary statistics derived from GWAS and LD data from a reference panel with penalization employed. In this study, the default panel (LDblocks = “ASN.hg19”) in the *lassosum* R package was employed as a reference. Subsequently, we selected the optimized elastic net tuning parameters (s and λ) through a process of validation. We selected s and λ values that resulted in the largest R^2 were chosen for subsequent investigation.

3.3 PRS-CS

PRS-CS employs a high-dimensional Bayesian regression framework and continuous shrinkage prior to determining the posterior effect sizes of SNPs through the use

of GWAS summary statistics and an external LD reference panel. This approach is described in detail in Ref. [19]. This study employed the PRS-CS pipeline with various global shrinkage parameters, ϕ , to obtain weights for each SNP. (1×10^{-5} , 1×10^{-4} , 1×10^{-3} , and auto). These weights were derived from external EAS populations from the 1000 Genomes Project. Subsequently, the EpPRS for POPs was subsequently computed as a weighted sum of SNP scores in the validation phase.

4. Evaluation of EpPRS

To assess the efficacy of each EpPRS, a model was created for each POP exposure, adjusting for sex, age, and the top three PCs. The evaluation of EpPRSs involved the selection of the most optimal running parameters within the Lasso, C + T, and PRS-CS methods (s and λ for Lasso; P value threshold for C + T SNP sets; ϕ for PRS-CS). This was done based on the R^2 values from the linear regression model, adjusted for age, sex, and PC1-3. The EpPRS with the highest value of R^2 among the three methods was designated as the final alternative in the field of POP genetics. We then explored the utility of EpPRSs as substitutes in scenarios where personal exposure data were unavailable.

4.1 Performance comparison across methods

The corresponding EpPRSs were derived from the summary statistics of 22 POPs in the validation phase using different methods based on the results of the training phase. The EpPRS with the highest value of R^2 for each POP was selected (**Table 1**). The C + T method yielded the most effective EpPRSs for four PBDEs and seven PCBs. Lasso performed best for the seven PCBs and OCPs, while PRS-CS exhibited superior performance for BDE100 and PCB77. These results highlight that the performance of each method varies depending on the specific trait, emphasizing the significance of evaluating multiple methods (**Figure 2A**).

4.2 Correlations of EpPRSs across exposures

To assess if EpPRSs can serve as an optional approach in instances when someone's exposure data is lacking, we examined the correlations between known exposures and EpPRSs. Most POPs exhibited positive relationships (r ranging from -0.23 for PCB157 and BDE153 to 0.73 for PCB170 and PCB180; **Figure 2B**). Similar but slightly lower correlations were observed between EpPRSs, with values ranging from -0.28 for PCB153-PRS & β -HCH-PRS to 0.49 for PCB138-PRS and PCB153-PRS (**Figure 2C**). Since EpPRSs capture only a small portion of the exposure variance, the pairwise correlations were generally lower (**Figure 2D**). Consequently, the correlations of POP measurement were validated to a satisfactory degree by EpPRSs according to consistent patterns. These data indicate that the EpPRSs are suitable candidates for exposure assessment in studies where direct measurement is impractical or biased.

4.3 Combined effect of POPs

Seeing the potential synergistic effects of combined exposures to POPs (PCBs, PBDEs, and OCPs) in leading to adverse effects, we aggregated the concentrations of

POPs	PRS generation			PRS evaluation
	Method	Tuning parameter	SNPs in EpPRS	R^2
PCBs	C + T	$r^2 = 0.001; P < 1 \times 10^{-3}$	361	0.0288
PCB77	PRS-CS	$\phi = 1 \times 10^{-4}$	770,688	0.0129
PCB81	C + T	$r^2 = 0.1; P < 1 \times 10^{-4}$	63	0.0471
PCB105	C + T	$r^2 = 0.01; P < 1 \times 10^{-4}$	62	0.0115
PCB114	Lasso	$s = 0.2; \lambda = 0.002069138$	2,712,774	0.0343
PCB118	C + T	$r^2 = 0.1; P < 1 \times 10^{-5}$	4	0.0431
PCB123	Lasso	$s = 0.2; \lambda = 0.001274275$	2,929,653	0.0397
PCB126	C + T	$r^2 = 0.1; P < 1 \times 10^{-4}$	58	0.0226
PCB138	Lasso	$s = 0.2; \lambda = 0.001$	3,006,480	0.0202
PCB153	C + T	$r^2 = 0.1; P < 1 \times 10^{-5}$	2	0.0353
PCB157	Lasso	$s = 1; \lambda = 0.001$	3,269,422	0.0226
PCB167	Lasso	$s = 1; \lambda = 0.1$	144,617	0.0411
PCB169	C + T	$r^2 = 0.01; P < 1 \times 10^{-4}$	77	0.0246
PCB170	C + T	$r^2 = 0.001; P < 1 \times 10^{-4}$	46	0.0285
PCB180	Lasso	$s = 1; \lambda = 0.078476$	379,390	0.0200
PCB189	Lasso	$s = 1; \lambda = 0.001$	3,267,800	0.0147
PBDEs	C + T	$r^2 = 0.1; P < 1 \times 10^{-3}$	515	0.0810
BDE47	C + T	$r^2 = 0.001; P < 1 \times 10^{-4}$	58	0.0623
BDE99	C + T	$r^2 = 0.01; P < 1 \times 10^{-5}$	9	0.0427
BDE100	PRS-CS	$\phi = 1 \times 10^{-4}$	770,688	0.0231
BDE153	C + T	$r^2 = 0.01; P < 1 \times 10^{-3}$	491	0.0500
BDE154	C + T	$r^2 = 0.1; P < 1 \times 10^{-3}$	488	0.0878
OCPs	C + T	$r^2 = 0.1; P < 1 \times 10^{-4}$	67	0.1395
β -HCH	Lasso	$s = 0.2; \lambda = 0.001$	3,010,774	0.1401
4,4'-DDE	Lasso	$s = 0.9; \lambda = 0.1$	83,547	0.0850

Table 1.
The best generated and evaluated PRS for each POP.

2 OCPs, 15 PCBs, and 5 PBDEs to assess their cumulative effects, respectively. Based on GWAS summary data (**Figure 2E**), we observed no systematic inflation of P values in accordance with the residual population structure ($\lambda_{\text{PCBs}} = 1.000$, $\lambda_{\text{PBDEs}} = 1.006$, $\lambda_{\text{OCPs}} = 1.011$, **Figure 2F**). In line with the performance of the 22 individual POPs, the C + T method outperformed PRS-CS and Lasso for the combined assessments of PCBs, PBDEs, and OCPs (**Figure 2G**). Similar patterns were found between exposure levels, EpPRSs, and exposure versus EpPRSs (**Figure 2H**).

4.4 Fungibility of EpPRSs for disease diagnosis compared to pollutant exposure

The measurement of the environmental pollutants concentrated in plasma is often invasive, time-consuming, and costly. We investigated whether EpPRSs could serve as

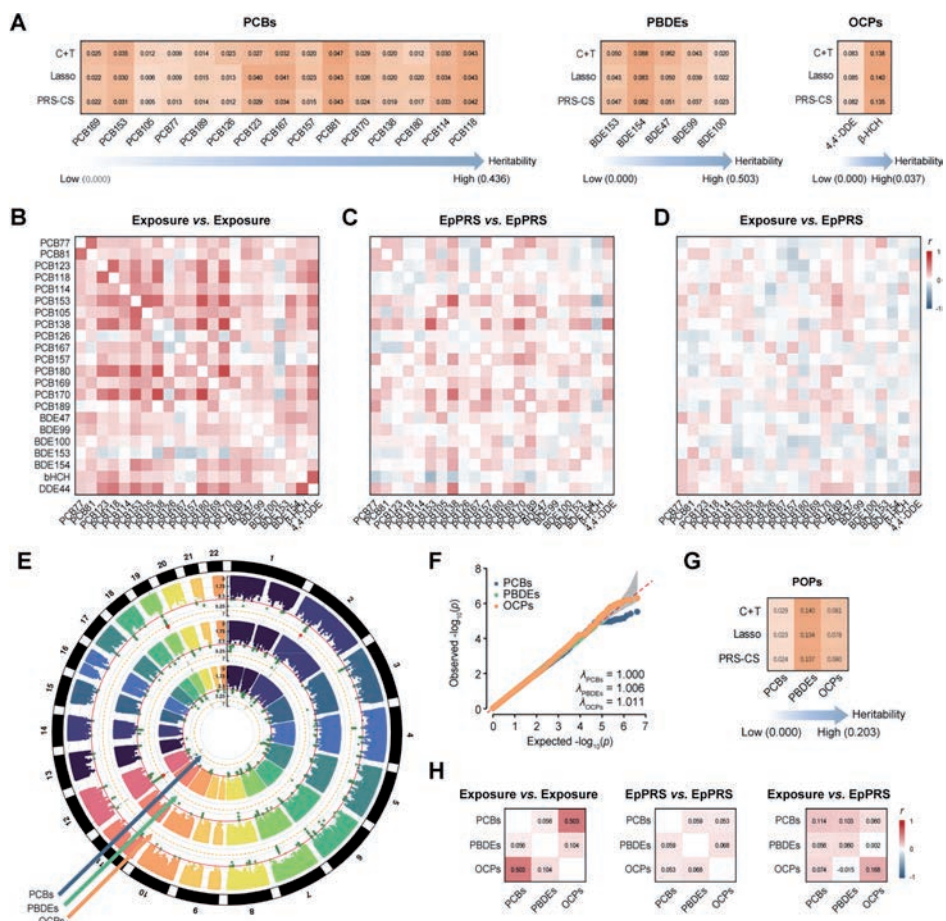


Figure 2.

Prediction performance and comparative analysis of EpPRSs across various methodologies and exposures. (A) The predictive performance of the three utilized PRS methods for 22 POPs was evaluated in the validation stage. The heatmap illustrates the relative predictive performance of each method across different POPs. The x-axis is arranged based on heritability in the training stage. We compare the pairwise correlation between (B) 22 continuous exposures in the validation stage, (C) EpPRSs in the validation stage, and (D) exposures (y-axis) and EpPRSs (x-axis). (E) Additionally, we present circular Manhattan plots of GWASs for the combined effect of PBDEs, PCBs, and OCPs. The outermost circle represents the chromosomal and SNP position on the x-axis, while from the periphery to the center, it indicates the GWASs of the OCPs, PBDEs, and PCBs, respectively. SNPs outside the red circular lines indicate $P < 5 \times 10^{-4}$. (F) The quantile-quantile plot illustrates genome-wide association with PCBs, PBDEs, and OCPs. Along with representation of genomic inflation factor λ for associations with POPs. (G) The predictive performance of the three utilized PRS methods of PCBs, PBDEs, and OCPs in the validation stage. The heatmap displays the relative predictive performance of each method across POPs. The x-axis is arranged based on heritability in the training stage. (H) Pairwise correlation between PCB, PBDE, and OCP exposures are compared in the validation stage, alongside EpPRSs and exposures on the y-axis and EpPRSs on the x-axis.

alternatives in the absence of individual exposure data. Owing to its high prevalence and significant links to environmental and genetic factors, T2DM was chosen for comparison. The predictive manifestations between personal phenotype and EpPRSs were usually measured using the area under the receiver operating characteristic curve (AUC). The DeLong test was utilized for AUC comparison. A non-significant DeLong test suggested that EpPRSs could effectively substitute exposure data. Most

EpPRSs were found to be reliable predictors for T2DM prediction, although not for PCB169, BDE47, BDE154, or 4,4'-DDE (**Figure 3**). Discrepancies between specific EpPRSs and original pollutant concentrations might stem from the restricted power of the exposure GWASs owing to small sample sizes.

4.5 Utilities of EpPRSs on outcomes

To advance investigate the diagnostic efficacy of EpPRSs, we contrasted the AUCs of a traditional disease-PRS model (T2DM-PRS) and a disease-EpPRS model (T2DM-PRS & EpPRS of POPs) (**Figure 4A**). Specifically, we first constructed a traditional T2DM-PRS using significant SNPs from reported T2DM GWAS data, applying the C + T method ($MAF \geq 0.05$, $r^2 < 0.1$, P value $< 5 \times 10^{-8}$) [20]. The T2DM-PRS was then utilized to assess the diagnostic effectiveness of T2DM. We then developed a disease-EpPRS model (T2DM-EpPRS) to investigate the combined EpPRS of POPs and the effect of T2DM-PRS. The diagnostic manifestations of the T2DM-EpPRS model for PCB126 (AUC = 0.694, 95% CI = 0.615–0.773, $P_{\text{delong test}} = 0.043$; **Figure 4B**) and for BDE153 (AUC = 0.703, 95% CI = 0.627–0.780, $P_{\text{delong test}} = 0.037$; **Figure 4C**) showed an improved accuracy for the prediction of T2DM compared to the T2DM-PRS alone (AUC = 0.625, 95% CI = 0.548–0.701).

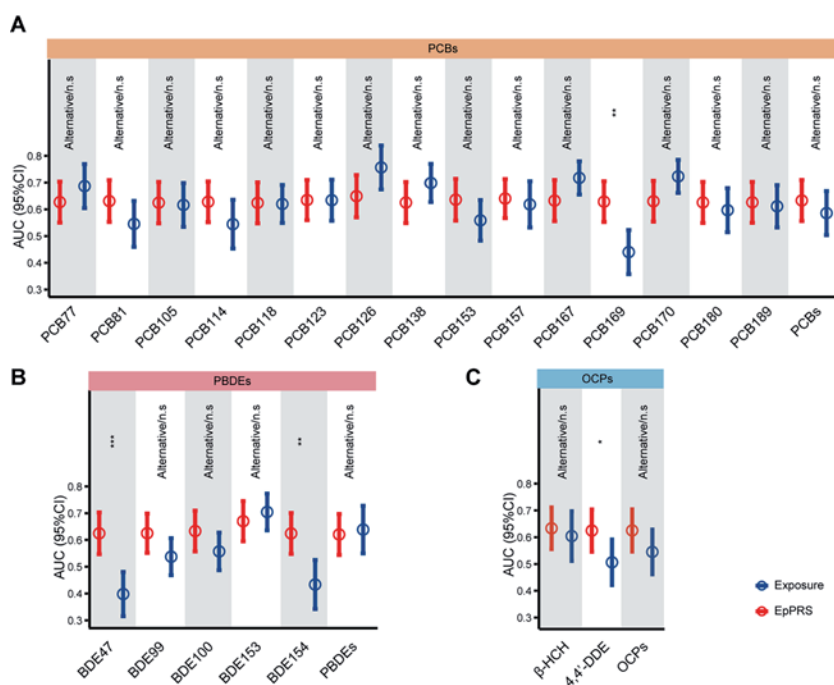


Figure 3. Comparing the predictive efficacy of exposure and EpPRSs for T2DM during the validation stage. The accompanying forest plots illustrate the AUC curves and 95% CI for EpPRS (depicted in red) and exposure (shown in blue) for (A) PCBs, (B) PBDEs, and (C) OCPs. The vertical extent of each bar displays the 95% confidence interval for the AUC, with the point at which the bar crosses the axis showing the AUC estimate. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Alternatives/n.s.: the AUCs between EpPRSs and exposures yielded insignificant results, indicating that EpPRS could serve as effective alternative methods for pollutants in disease prediction.

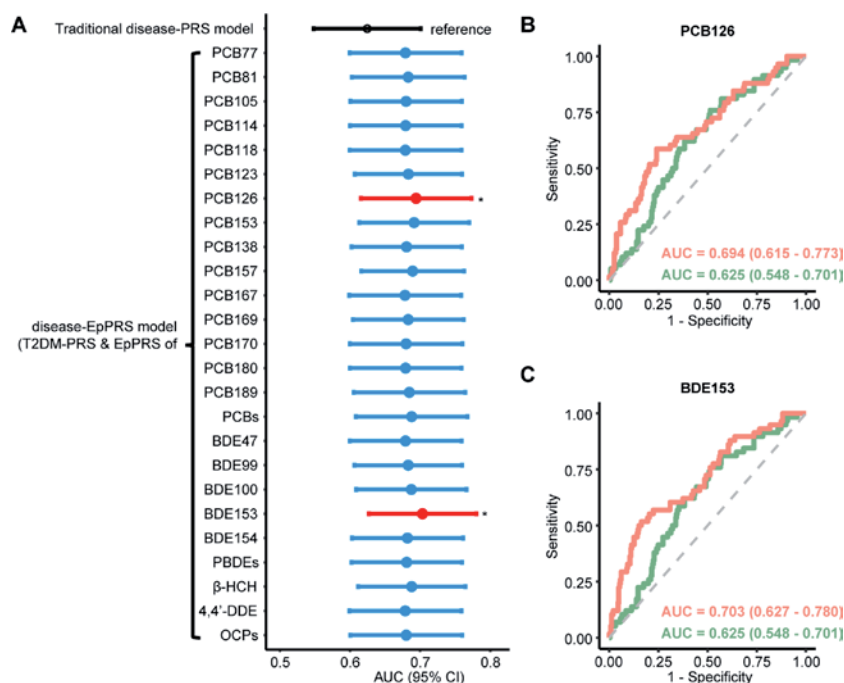


Figure 4.

Assessing the diagnostic efficacy of two models for the prediction of T2DM: a traditional disease-PRS and a disease-EpPRS model. (A) The forest plot illustrates the performance of the two models in predicting T2DM. In comparison to the traditional disease-PRS model, the disease-EpPRS model demonstrates a statistically significant improvement in the AUC curves and 95% CI, which indicates enhanced diagnostic performance for T2DM. (B) ROC curves for the diagnostic accuracy of the traditional disease-PRS model and disease-EpPRS of the PCB126 model for T2DM. (C) ROC curves illustrate the performance of the traditional disease-PRS and disease-EpPRS of the BDE153 models for T2DM. The efficacy of the traditional disease-PRS model was shown in green ROC curve, while the efficacy of the disease-EpPRS model was illustrated in the orange ROC curve.

5. Application of EpPRS

5.1 AOP framework application in EpPRS biological pathways

5.1.1 Identification of potential molecular initiating events and key events

In order to further explore the underlying mechanisms linking POP exposure and T2DM, the AOP framework was established [7]. To identify specific POPs serving as possible MIEs contributing to T2DM, we focused on PCB126 and BDE153, due to their significant enhancement of T2DM prediction through EpPRSs. A total of 240 associated with contaminant genes were identified from the SNPs annotated in the EpPRSs of PCB126 and BDE153. Through cross-referencing the GeneCards database, we identified 2626 genes closely relevant to T2DM. (GIFtS ≥ 10 and relevance score ≥ 5 ; <https://www.genecards.org/>) [21]. As illustrated in **Figures 5**, 47 pollution-disease-related genes were found to overlap between the two databases.

For these, which may be considered potential MIEs, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis, which yielded 2713 phenotypes associated with T2DM (2690 GO terms and 23 pathways). In addition, utilizing the Comparative Toxicogenomics

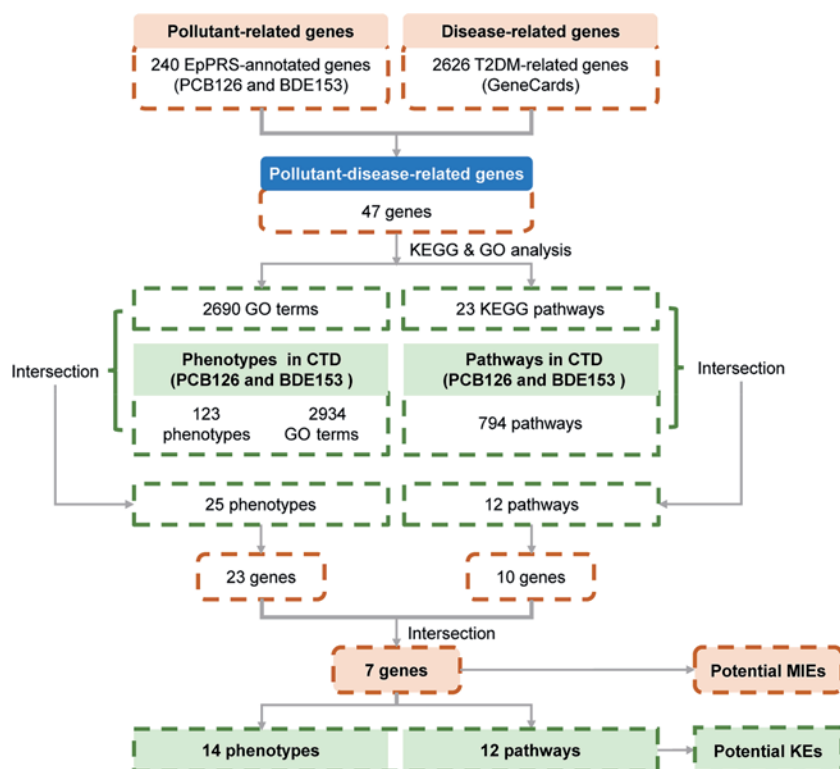


Figure 5.
 The flow diagram illustrates the construction of an AOP framework.

Database (CTD, <https://ctdbase.org/>) [22], we identified 3851 phenotypes relevant to T2DM (123 phenotypes, 2934 GO terms, and 794 pathways). We found seven coregulated genes (KCNQ1, HMGB1, AKT1, UGT1A9, CACNA1C, UGT1A7, and ABCC1) among the overlapping phenotypes for T2DM as potential MIEs. Potential key events (KEs) were identified for the phenotypes (14 phenotypes and 12 pathways) regulated by these seven genes.

5.1.2 Construction and assessment of the AOP framework

The AOP framework was built upon direct MIEs and the interrelationships between adverse outcomes relevant to risk assessments across various organizational levels in organisms. Unlike previous studies that often had isolated foci (*in vitro/in vivo*), our AOP framework standardized and modularized events across molecular, cellular, tissue, organ, and ultimately organismal, as well as population levels. This approach emphasizes logical inference for the determination of the contextual link between POP exposure and disease occurrence. We utilized a hierarchical diagram of the GO terms (<https://www.ebi.ac.uk/QuickGO/>) to categorize the pathways and phenotypes, which were categorized into three groups: system, cellular, and KEGG pathways.

For assessing this AOP framework, we applied the weight of evidence (WoE) to assess the data, identify each toxic exposure event, and highlight the uniformity of

similar toxic events across different toxic procedures. WoE was also used to link the series of various toxic events, ensuring logical connections between toxic events. According to the handbook from the Organization for Economic Co-operation and Development (OECD) handbook, AOPs were categorized as low, moderate, or high confidence based on the evidence's robustness.

Within the resulting AOP framework, we divided cellular processes into six groups. As illustrated in **Figure 6A**, the AOP framework was developed through comprehensive data collection, systematic evaluation of the research findings, and assessment of biological plausibility using the WoE method. Our AOP framework indicated that the combined effects of genetic and environmental stressors from PCB126 and BDE153 influenced the expression of seven potential MIEs. These stressors included cellular processes, particularly homeostatic and metabolic processes, via the cGMP-PKG signaling pathway and the sphingolipid signaling pathway. At a systemic level, alterations in these cellular processes resulted in liver damage, a critical organ for glucose metabolism, and ultimately to T2DM.

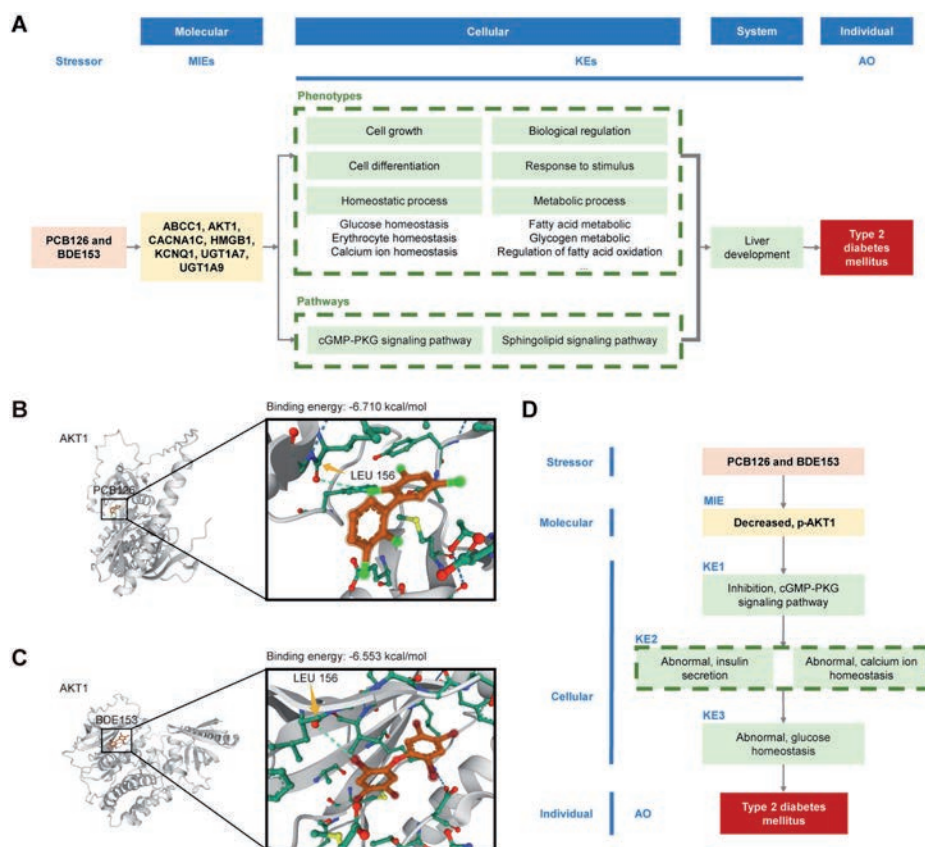


Figure 6.

A framework for the analysis of the underlying mechanisms of action of POPs in the development of T2DM, from the alteration of gene expression to the manifestation of the disease. (A) An AOP framework connecting PCB126 and BDE153 susceptibility genes to T2DM. (B) Molecular docking analysis of PCB126 and AKT1. (C) Molecular docking analysis of BDE153 and AKT1. The cyan dotted line illustrates the formation of a hydrogen bond. (D) AKT1-mediated AOP framework of T2DM.

5.1.3 Construction and assessment of the AKT1-mediated AOP framework

It is interesting to note that AKT serine/threonine kinase 1 (AKT1) was found to participate in most of the potential KEs during our study of the seven potential MIEs. To evaluate the binding affinity of PCB126 and BDE153 to AKT1, we conducted *in silico* molecular docking to describe the structural interactions. The structure of AKT1 was obtained from the AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/>, UniProt ID: P31749). The 3D structures of POPs were created using PubChem CID (<https://pubchem.ncbi.nlm.nih.gov/>), and PyMOL visualized the structural interactions between AKT1 and POPs. The binding patterns revealed that both BDE153 and PCB126 formed a hydrogen bond with LEU-156 of AKT1, with predicted binding energies of -6.710 kcal/mol for PCB126 (**Figure 6B**) and -6.553 kcal/mol for BDE153 (**Figure 6C**), which indicates stable binding.

Consequently, we identified AKT1 as a possible MIE for developing the AKT1-mediated AOP framework (**Figure 6D**). In this framework, we noted that the combined environmental and genetic stressors inhibited AKT1 phosphorylation. Subsequently, the downregulation of AKT1 phosphorylation disrupted the cGMP-PKG signaling pathway, leading to abnormalities in insulin secretion and calcium ion homeostasis. This disruption resulted in impaired glucose homeostasis at the organ-system level, resulting in T2DM. Evaluation of the WoE for the AOP following handbook from OECD indicated high confidence in this framework, suggesting its applicability in assessing the impact of POP exposure on T2DM.

5.2 Mendelian randomization analysis

Apart from employing genes annotated by EpPRS to explore disease toxicity pathways, POP GWAS SNPs can be used as instrumental variables (IVs) to investigate the causal relationship between POPs and disease. Hence, we suggest making use of Mendelian randomization (MR) analysis, a causal association study approach that necessitates summary statistics data in lieu of raw genotype data, to investigate the connection between POPs and cancer risk, as an extension of EpPRS.

Furthermore, POP GWAS SNPs can be employed for causal inference. Using the C + T method-based SNPs as IVs, we performed a two-sample MR analysis to explore the causal relationship between 11 POPs and the risk of 11 cancers in BioBank Japan (<https://jenger.riken.jp/en/>) [23]. The analyses were conducted using *TwoSampleMR* R package (version 0.5.6). It is noteworthy that the carcinogenic impact of POPs was mainly displayed in digestive system cancers, including BDE154 and pancreatic cancer (OR = 1.03, $P = 0.016$), PCB81 and colorectal cancer (OR = 1.02, $P = 0.017$), PCB105 and gastric cancer (OR = 1.02, $P = 0.018$), and PCB126 and biliary tract cancer (odds ratio (OR) = 1.08, $P = 0.047$). These findings highlight the utility of EpPRS in extending beyond T2DM to explore associations with other health outcomes, offering a robust tool for understanding the broader impacts of environmental pollutants on human health.

By integrating genetic data and exposomic information, the EpPRS framework not only advances our understanding of the genetic and environmental interplay in disease risk but also provides a valuable resource for public health strategies aimed at mitigating the adverse effects of environmental pollutants. The application of EpPRS in different contexts demonstrates its versatility and potential for significant contributions to environmental health research.

6. Discussion

In this chapter, we established EpPRSs using environmental genomics data and examined their feasibility as genetic substitutes for contaminant exposure. Taking POPs and T2DM as an illustration, we demonstrated that EpPRS-annotated genes allow for the rapid, efficient, and accurate investigation of the toxic hazards posed by chemical substances. Furthermore, by employing the AOP framework, we revealed the mechanism of action through which these chemical substances impact disease development.

In a previous study, Ma and colleagues demonstrated that an exposure PRS could be used to improve disease prediction models, which could serve as a proxy for exposures that are hard to measure. While our study shares conceptual similarities, it uniquely focuses more on environmental contaminant-susceptible SNPs, an area that remains underexplored in environmental genomics. However, it is important to note that EpPRS is solely capable of quantifying the heritable component of exposure variability associated with genetically determined variants at the time of birth. It does not account for early variants or lifetime environment exposure [24]. In the case of pollutants with impacts that show high heritability, as indicated by large-sample GWASs, EpPRS can serve as a more accurate replacement for direct concentration measurement [8]. Of note, the methodology discussed in this chapter can be expanded to other environmental pollutants.

Nevertheless, this study has certain limitations. Our subjects were exclusively from the Asian population, and we did not employ the novel PRS-CSx method, which could boost the clinical applicability of PRS across diverse ethnic groups [25]. Additionally, our results indicated that generating EpPRS with the best predictive performance does not follow a one-size-fits-all method. Instead, a wide variety of methodological options must be considered to optimize predictions.

The application of EpPRSs extends beyond model prediction and AOP framework construction to include MR analysis. Additionally, mediation analysis can also be employed to investigate polygenic pleiotropy. Furthermore, phenome-wide explorations can identify clinical phenotypes potentially associated with environmental exposure, offering a comprehensive approach to understanding the health impacts of environmental pollutants.

7. Conclusion

This chapter introduced the EpPRS as a valuable proxy for evaluating exposure to environmental contaminants, especially in studies where direct measurements are impractical or biased. Our findings provide robust evidence supporting the use of EpPRS in disease risk assessment related to environmental pollutants, thereby providing information for strategies of regional contaminant control. We anticipate that this research will significantly contribute to decision-making processes aimed at mitigating the health risks associated with environmental pollution.

Acknowledgements

This project was funded by the Natural Science Foundation of Jiangsu Province (BK20230003) and the Science and Technology Innovation Team (Tianshan Innovation Team) Project of Xinjiang Uighur Autonomous Region (2022TSYCTD0013).

Conflict of interest

The authors declare no conflict of interest.

Author details

Mulong Du^{1*}, Junyi Xin^{2*}, Silu Chen³ and Meilin Wang^{3†}

1 Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China

2 Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China

3 Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, China

*Address all correspondence to: drdumulong@njmu.edu.cn
and junyixin@njmu.edu.cn

† Senior author who supervised this work.

IntechOpen

© 2024 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Avila BS, Mendoza DP, Ramirez A, Penuela GA. Occurrence and distribution of persistent organic pollutants (POPs) in the atmosphere of the Andean city of Medellin, Colombia. *Chemosphere*. 2022;**307**(Pt 1):135648
- [2] Castro-Jimenez J, Banaru D, Chen CT, Jimenez B, Munoz-Arnanz J, Deviller G, et al. Persistent organic pollutants burden, trophic magnification and risk in a pelagic food web from coastal NW Mediterranean Sea. *Environmental Science and Technology*. 2021;**55**(14):9557-9568
- [3] Wang S, Ding G, Liu Y, Dou Z, Chen H, Ya M, et al. Legacy and emerging persistent organic pollutants in the marginal seas of China: Occurrence and phase partitioning. *Science of the Total Environment*. 2022;**827**:154274
- [4] Alava JJ, Cheung WWL, Ross PS, Sumaila UR. Climate change-contaminant interactions in marine food webs: Toward a conceptual framework. *Global Change Biology*. 2017;**23**(10):3984-4001
- [5] Jones KC. Persistent organic pollutants (POPs) and related chemicals in the global environment: Some personal reflections. *Environmental Science and Technology*. 2021;**55**(14):9400-9412
- [6] Papadopoulou E, Haug LS, Sakhi AK, Andrusaityte S, Basagana X, Brantsaeter AL, et al. Diet as a source of exposure to environmental contaminants for pregnant women and children from six European countries. *Environmental Health Perspectives*. 2019;**127**(10):107005
- [7] Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, et al. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry*. 2010;**29**(3):730-741
- [8] Ma Y, Patil S, Zhou X, Mukherjee B, Fritsche LG. ExPRSweb: An online repository with polygenic risk scores for common health-related exposures. *American Journal of Human Genetics*. 2022;**109**(10):1742-1760
- [9] Li S, Xu S, Chen Y, Zhou J, Ben S, Guo M, et al. Metal exposure promotes colorectal tumorigenesis via the aberrant N(6)-Methyladenosine modification of ATP13A3. *Environmental Science & Technology*. 2023;**57**(7):2864-2876
- [10] Parajuli RP, Goodrich JM, Chan LHM, Ayotte P, Lemire M, Hegele RA, et al. Genetic polymorphisms are associated with exposure biomarkers for metals and persistent organic pollutants among Inuit from the Inuvialuit settlement region, Canada. *Science of the Total Environment*. 2018;**634**:569-578
- [11] Parajuli RP, Goodrich JM, Chan HM, Lemire M, Ayotte P, Hegele RA, et al. Variation in biomarker levels of metals, persistent organic pollutants, and omega-3 fatty acids in association with genetic polymorphisms among Inuit in Nunavik, Canada. *Environmental Research*. 2021;**200**:111393
- [12] Kelly BC, Ikonomou MG, Blair JD, Morin AE, Gobas FA. Food web-specific biomagnification of persistent organic pollutants. *Science*. 2007;**317**(5835):236-239
- [13] American Diabetes A. Standards of medical care in diabetes--2010. *Diabetes Care*. 2010;**33**(Suppl_1):S11-S61

- [14] Wang L, Peng W, Zhao Z, Zhang M, Shi Z, Song Z, et al. Prevalence and treatment of diabetes in China, 2013-2018. *JAMA*. 2021;**326**(24):2498-2506
- [15] Wang M, Gu D, Du M, Xu Z, Zhang S, Zhu L, et al. Common genetic variation in ETV6 is associated with colorectal cancer susceptibility. *Nature Communications*. 2016;**7**:11478
- [16] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*. 2011;**88**(1):76-82
- [17] Choi SW, Mak TS, O'Reilly PF. Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*. 2020;**15**(9):2759-2772
- [18] Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genetic Epidemiology*. 2017;**41**(6):469-480
- [19] Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*. 2019;**10**(1):1776
- [20] Suzuki K, Akiyama M, Ishigaki K, Kanai M, Hosoe J, Shojima N, et al. Identification of 28 new susceptibility loci for type 2 diabetes in the Japanese population. *Nature Genetics*. 2019;**51**(3):379-386
- [21] Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics*. 2016;**54**:1 30 31-31 30 33
- [22] Davis AP, Wiegiers TC, Johnson RJ, Sciaky D, Wiegiers J, Mattingly CJ. Comparative toxicogenomics database (CTD): Update 2023. *Nucleic Acids Research*. 2023;**51**(D1):D1257-D1262
- [23] Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nature Genetics*. 2020;**52**(7):669-679
- [24] Shi J, Swanson SA, Kraft P, Rosner B, De Vivo I, Hernan MA. Instrumental variable estimation for a time-varying treatment and a time-to-event outcome via structural nested cumulative failure time models. *BMC Medical Research Methodology*. 2021;**21**(1):258
- [25] Ruan Y, Lin YF, Feng YA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*. 2022;**54**(5):573-580

Genetic Changes, Epigenetic Modifications and Antibiotic Resistance Caused by Bacterial Infection

Mohammad Hashemzadeh and Aram Asareh Zadegan Dezfuli

Abstract

Genetic elements play a crucial role in spreading recombinant strains with new traits. Bacteria share genetic elements and transfer them to one another through various means. The transferred genetic material often benefits the recipient bacteria, providing advantages such as antibiotic resistance, tolerance to foreign substances, or the ability to utilize new metabolites. This utility is essential for bacteria to thrive in different environmental conditions and in the human body. This study investigates the roles of genetic elements in both Gram-negative and Gram-positive bacteria.

Keywords: genetic modification, antibiotic-resistant, infection disease, gram-positive bacteria, gram-negative bacteria

1. Introduction

The field of microbial genetics has its roots in the 1950s and is based on the emergence of genetic variants in microorganisms and the deliberate induction of genetic variations through chemical and physical mutagens. Microorganisms also reproduce quickly, which facilitates genetic experimentation [1]. Researchers have used natural methods, such as transformation, conjugation, and virus-mediated transduction, to create temporary or more permanent partial diploidy in bacteria. This has allowed them to study how genetic material complements each other and how genetic recombination occurs between different variants [2]. By using these methods, scientists have been able to assign individual mutations to specific genes and establish genetic maps [3]. The same principles apply to the study of bacteriophage genetics when different phage mutants infect host bacteria simultaneously. In classical microbial genetic studies, mutants induced by specific mutagens—resulting in nucleotide substitutions and missense or nonsense mutations—have been extensively used. This has been possible thanks to the understanding of how specific mutagens work. One common assumption in this field has been that spontaneous genetic variants

are usually substitution mutants or, more broadly, point mutants [2]. In the late 1960s, microbial geneticists started concentrating on a specific kind of spontaneous mutation that has polar effects, which are caused by the insertion of mobile genetic elements [4]. Further research on the movement of these elements showed that not all microbial recombination activities require homologous recombination between identical DNA sequences [5, 6]. Site-specific recombination, such as the integration of proviral genomes into the bacterial host chromosome, was discovered. It was revealed that recombination can occur in DNA sequences. Transposition and site-specific recombination result in deletions and DNA inversions [7]. These processes can also bring together chromosomal DNA segments, serving as natural gene carriers. In this chapter, genetic elements and their role in bacterial genome changes, and the transfer of different genes, including antibiotic resistance, are discussed.

2. The role of genes in the creation of genetic distinctions

Hereditary variety is essential for organic advancement. Populaces with assorted parental shapes and hereditary variations are the premise for common choice, affected by the organisms' living conditions. Studies show that genetic variants are caused by mistakes during DNA replication that cause large changes in the original sequence [8]. The next hypothesis is the evolutionary genes that show the biological evolution of the population of organisms. In general, there are two evolutionary gene models [9]. One of these species produces genetic variation, while the other species control genetic variation. Bacteria have different forms in terms of adaptation to the environment, nutritional deficiency, and inappropriate conditions of temperature, pressure, and oxygen [10]. Also, some bacterial species have become the normal flora of the body and have a peaceful coexistence with many of their hosts. Of course, this condition exists as long as the host has a good safety situation. Today, good studies of gene evolution in bacteria are in access [11, 12].

3. Discover natural strategies for creating genetic variations

The concept of mutation in molecular genetics differs from classical genetics. In molecular genetics, any change in the nucleotide sequence of the genome is considered a mutation, regardless of whether it causes a change in the organism's physical characteristics. Molecular mechanisms that cause DNA sequence alterations do not necessarily result in immediate phenotypic changes [13]. Many nucleotide changes have no immediate effect on the organism's phenotype. While few DNA sequence alterations offer a selective advantage to the organism, many more cause a selective disadvantage or lethality [14]. However, genetic changes do not occur spontaneously in bacteria only in response to a different environmental condition. There are three genetic strategies for creating genetic changes in bacteria. (1) DNA sequence changes, such as nucleotide substitution, deletion, insertion, and disruption of the sequence of nucleotides, (2) Rearrangement of parts of the DNA sequence through recombination displacement, (3) Horizontal gene transfer, which leads to the acquisition of "foreign" DNA sequences. In **Figure 1**, a comprehensive overview of the molecular evolution elements of prokaryotic microorganisms is displayed [15].

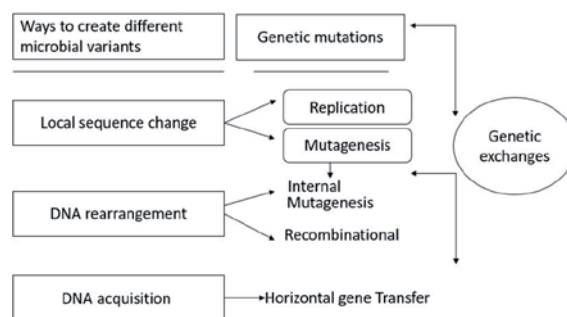


Figure 1.
 Comprehensive overview of the molecular evolution elements of prokaryotic microorganisms.

There are three main strategies for creating genetic variations, and they each have different effects. When it comes to their impact on evolution, different mechanisms within the same strategy can replace each other, but not mechanisms from a different strategy. This means that bacteria that are well-suited for evolution are likely to have at least one, and preferably several, systems for each of the three different evolutionary strategies [16].

4. Local sequence change

Local sequence changes could theoretically be individual steps in developing new biological functions, but there is no general evidence for a specific direction of these steps. The creation of a completely new biological function through mutations maybe extremely rare unless natural selection can act on a basic function controlled by the DNA sequence. The situation may change as biological functions are gradually improved and natural selection begins to exert its influence. This applies to different genes. Researchers use ribosomal RNA to estimate evolutionary distances in the place of gene sequence changes. In this way, harmful mutations in action-oriented genes are eliminated. Also, changes in the sequence of the DNA sequence affect its structure [17].

5. DNA rearrangement

Gene conversion can also occur through the process of intragenomic reshuffling of DNA sequences, where related sequences within the genome undergo recombination. Another important effect of DNA rearrangement is the creation of new combinations of functional capacities through the fusion of different functional domains. Genomics research increasingly reveals that different functional genes often share related sequence motifs and functional domains [18]. Occasional DNA reshuffling is a significant source for the creation of these fusion genes. The reassortment of expression control signals with different reading frames used in protein production also occurs through DNA rearrangement. Additionally, DNA rearrangements can lead to the duplication of DNA segments, providing substrates for further evolution. Conversely, deletion of DNA segments can also result from DNA rearrangement processes, helping to remove unnecessary sequences from the genome through natural selection [19].

6. DNA acquisition

Antibiotic-resistance genes have been widely studied, and it has been found that microorganisms can share the evolutionary success of others through DNA acquisition. This efficient strategy can quickly result in new capabilities, particularly for accessory genes, and can enable microorganisms to inhabit new environments. This strategy has a significant evolutionary impact. Additionally, the acquisition strategy can also influence the functions of essential housekeeping genes through the transfer of DNA segments [20].

7. Recombinant systems

Many of the molecular functions of site-specific recombination systems of bacteria have been extensively studied. Microbial populations may evolve different DNA arrangements as a result of some of these systems. A significant number of secondary recombination sites can be created in DNA by enzymes that invert DNA, each with its recombination frequency. For an individual reaction, the results of such recombination are not predictable statistically. Although hybrid locations differ from the consensus, they are relatively stable, displaying local sequence variations in addition to the consensus [21]. Scientists can interpret DNA inversions involving secondary sites of recombination as errors caused by enzyme systems or as evidence of evolution. Site-specific DNA inversions largely provide novel DNA sequence fusions at low frequencies that maybe useful under natural selection, highlighting the evolutionary function of site-specific DNA inversions [22]. In flip-flops, DNA inversions are more frequently reproducible, possibly as a result of the evolution of the gene involved. When none of the possible arrangements of a DNA segment exhibit a strong selective disadvantage, a segment undergoing frequent reproducible inversion can be tolerated [23].

8. Transferable elements

There is an impressive number of molecular elements known to exist in prokaryotes, including insertion sequences (IS). These components exhibit different DNA rearrangement capabilities that are determined by the individual selection criteria for each component [24]. Some IS elements adhere strictly to consensus sequences for their insertion targets, while others adhere to more intricate rules and may have a preference for specific regions of DNA where they can insert into various DNA sequences. Transpositions of DNA do not lead to gene fusion right away [25]. However, the removal of an IS element at a later stage can cause gene fusion. IS elements can impact genome evolution by offering sequence homologies in different DNA locations, resulting in DNA rearrangements through homologous recombination. These changes, together with transpositional DNA rearrangements, play a role in both the joining and separating of chromosomal DNA segments to and from natural gene carriers, which is important for gene acquisition strategies. In general, ARE elements control their transposition activities to ensure acceptable rates of DNA rearrangements for maintaining a specific microbial strain [26]. However, certain IS elements can cause very high rates of various DNA rearrangements due to specific structural arrangements at the location of IS elements in DNA molecules (dimeric IS

structures). This can result in a sudden increase in transpositions within a subset of the bacterial strain containing the IS element in question [27].

9. DNA repair systems

The majority of genetic rearrangements within a genome and mechanisms for transferring DNA horizontally are likely facilitated by genetic processes. However, when it comes to genetic variation caused by small changes in DNA sequence, the circumstances may be different. One of the primary factors contributing to such changes is the inherent instability of nucleotides' structure, which can result in mutations or errors during DNA replication [28]. Furthermore, external factors known as environmental mutagens can interact with DNA and induce alterations in the local sequence [29]. While generating genetic diversity through small changes in DNA sequence can benefit from these natural mutation sources, organisms must be careful to control the frequency of such alterations to manageable levels. This control is achieved through DNA repair mechanisms, essential for maintaining a certain level of genetic stability. Therefore, these mechanisms are considered evolutionary genes that regulate the rate of genetic variation [30].

10. Restriction-modification systems

Restriction enzymes play a significant role in reducing the uptake of foreign DNA by bacteria, which effectively lowers the acquisition of foreign gene functions to manageable levels. When foreign DNA molecules enter recipient bacteria, restriction enzymes cut them into fragments, which are then subject to degradation [31]. However, these fragments can also be recombined and permanently integrated into the recipient's genome if given the opportunity. In this way, bacterial restriction-modification systems demonstrate the operational genes of evolution. On one hand, they limit DNA acquisition to minimal levels, while on the other hand, they promote DNA acquisition in small increments [32].

11. Site-specific recombination

When discussing instances of genetic evolution, it is crucial to recognize that nature also utilizes factors other than genes to generate genetic variations. Both intrinsic structural characteristics of biological macromolecules and external influences, like those induced by mutagens interacting with DNA, play a crucial role [33]. This concept extends beyond just local sequence modifications, impacting various methods of creating genetic diversity. It is suggested that occasional interactions between site-specific recombination enzymes and secondary recombination sites could lead to temporary structural alterations in either the DNA substrate or the enzyme protein, potentially signifying significance. Similarly, structural instabilities could be implicated in the transposition of variants into consensus target sequences [34]. While acknowledging the potential difficulty of experimentally investigating this hypothesis using biophysical or biochemical approaches due to the inefficacy of these reactions and the fleeting nature of structural variants in biological macromolecules, it is apparent that structural

factors can also influence genetic variations. Another nongenetic component affecting the generation of genetic diversity is the random interactions between enzymes and their substrate, like recombination sites on DNA molecules [35]. These chance encounters also play a role in DNA acquisition strategies, such as in transformation, transduction, and conjugation processes [36]. Despite numerous efforts to determine whether bacteria possess mechanisms to identify evolutionary needs and respond by increasing mutagenesis or producing appropriate genetic variations, definitive answers remain elusive. Generally, the generation of genetic variations occurs independently of specific functional requirements related to particular environmental conditions. However, there may be exceptions to this principle, which can be explained by the adage ‘no rule without exception [37].

12. Evolution of evolutionarily

The components involved in biological evolution are likely to undergo evolution themselves. However, since the outcomes of evolution, such as genes, do not directly undergo natural selection within individual cells of microbial populations, their improvement occurs at the population level through their mutagenic impact on genes that can be directly selected [38]. This is referred to as second-order selection and has led to a refinement of evolutionary functions over time. This refinement may include aspects such as the interaction of recombination enzymes with different DNA sequences and the regulation of the transposition activity of mobile genetic elements [39]. It is also likely that elements within genomes have their evolutionary history, particularly nucleotide sequences that influence regional differences in how genetic variations are generated. Molecular genetics strategies and methodologies allow us to experimentally explore these questions and measure the contribution of specific mechanisms of genetic variation to overall mutagenesis [39].

13. Mobile genetic elements and antimicrobial resistance

Antibiotic-resistant bacteria are a significant contributor to infections acquired in healthcare settings globally, and resistance has also been observed in community-acquired infections. Illness, mortality, and healthcare costs are substantially increased by infections caused by multidrug-resistant organisms [40]. Studies at the molecular level have indicated that widespread multidrug resistance often results from the acquisition of preexisting determinants followed by their amplification in response to selective pressure. The movement, accumulation, and dissemination of resistance genes are predominantly facilitated by mobile genetic elements (MGEs), which are entities that facilitate the transfer of DNA within a cell (e.g., between a chromosome and a plasmid or between two plasmids) as well as between cells. Insertion sequences (IS) and transposons (Tn) are discrete segments of DNA capable of relocating themselves (along with associated resistance genes) to new positions almost indiscriminately within a single cell or between distinct DNA molecules. On the other hand, integrons (In) utilize specific recombination mechanisms to transfer resistance genes to defined sites. MGEs are often present in multiple copies in different regions of a genome, allowing for the exchange of genetic sequences through homologous recombination [41]. Genetic exchange between cells can be facilitated by various

mechanisms such as conjugation/mobilization (mediated by plasmids and integrative conjugative elements [ICE]), transduction (assisted by bacteriophages), and transformation (the uptake of extracellular DNA). The interplay between different MGEs plays a significant role in the rapid evolution of multiple multidrug-resistant pathogens in response to antibiotic therapy [42]. While it is not possible to cover all MGEs involved in resistance across all bacterial species in a single article, this discussion primarily focuses on the most noteworthy elements in Gram-negative and Gram-positive bacterial species that are of major clinical concern. Included in this list are *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter spp.* (known collectively as the ESKAPE group) as well as *Escherichia coli* (comprising the ESKAPEE group) [43]. However, the types of MGEs discussed are broadly applicable to numerous other bacterial groups. The objective is to highlight the diversity and key features of MGEs associated with resistance in these organisms while also referencing more detailed reviews for a deeper analysis of specific elements or biological processes [44].

Figure 2: The diagram illustrates various mobile genetic elements (MGEs) and the mechanisms involved in the transmission of antibiotic-resistance genes within and between cells.

13.1 IS and transposons

Insertion Sequences (IS) are usually small mobile genetic elements that typically carry little more than one (sometimes two) transposase (*tnp*) genes. Their characteristics have been reviewed several times. IS can be categorized based on

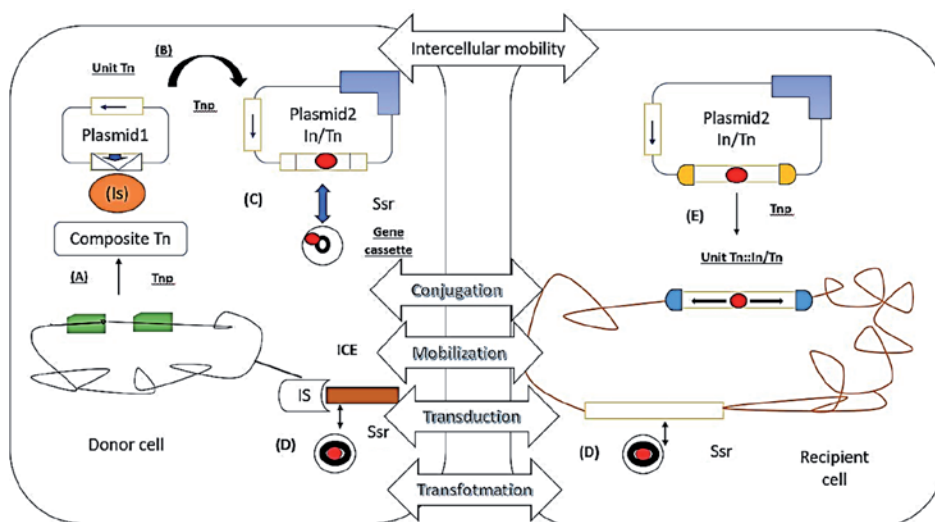


Figure 2. The diagram demonstrates the movement of antibiotic-resistance genes through mobile genetic elements in cells. Two cells are shown, one as a donor and the other as a recipient. The donor cell, highlighted in blue, contains two plasmids with various genes and mobile genetic elements of different colors. Intracellular processes are indicated by thin black arrows, with processes mediated by a transposase protein marked as Tnp and those by a site-specific recombinase protein as Ssr. Intercellular transfer is represented by thick green arrows. The illustration shows how resistance genes linked to different mobile genetic elements can move within and between DNA molecules, like from the chromosome to a plasmid or between plasmids. Additionally, the diagram displays the movement of gene cassettes, integrative and conjugative elements, and the transfer of resistance genes from incoming plasmids to the chromosome or other plasmids in the recipient cell.

active site motifs in Tnp, designated by key amino acids that come together in the active site [45]. These motifs are most commonly DDE (Asp, Asp, and Glu), but can also be DEDD or HUH (two His residues separated by a large hydrophobic amino acid). Additionally, IS can be classified based on whether transposition is conservative (cut-and-paste mechanism) or replicative. Replicative transposition can occur by a copy-and-paste mechanism (the IS top replicated to join the donor and recipient in a cointegrate, which is then resolved to give the original donor plus the recipient with the IS) or a copy-out-paste-in mechanism. Unit transposons have traditionally been considered as elements larger than IS, bounded by IR rather than by a pair of IS. They include a transposase gene and an internal “passenger” gene, which may encode antibiotic resistance [46]. The distinction between IS and Tn is becoming more complex, as there are now examples of relatives of well-known IS carrying passengers (transporter IS), which may have a Tn name/number, and cryptic relatives of transposons without any passengers, which may be given an IS name. Insertion sequences (IS) are mobile genetic elements that overlap the two transposase (tnp) genes. IS classification is usually classified based on movement, which includes cutting and pasting or repetitive movement. In the replication transfer mechanism, IS attaches a donor to a recipient or removes a copy [47]. Transposons are elements larger than IS that are controlled by IR. Transposons contain a transposase gene and antibiotic-resistance genes. Sometimes it becomes difficult to distinguish between IS and Tn because they have similar functions. Antibiotic-resistance genes are often associated with Tn3 family transposons. Members of the extended Tn3 family are typically characterized by a terminal IR of ~38 bp, designated IRL and IRR relative to the direction of transcription of the tnpA transposase gene [48]. Tn3 family transposons contain resolvase genes tnpR and a cleavage site (res). Translocation occurs *via* a multiplex mechanism in which TnpA generates a concentric structure containing repeated copies of the transposon that separates the original donor and acceptor molecules [49]. So, by direct recombination, the targeting site produced by TnpR is converted into molecules that contain one copy of the transposon. The translocation creates a 5 bp TSD. **Tables 1 and 2** show insertion sequences (IS) and compound transposons associated with resistance genes in Gram-negative and Gram-positive bacteria.

13.2 Gene cassette and integrons

A gene cassette is a small movable element, usually sized between 0.5 and 1 kilobase pairs, containing a single gene (sometimes two), typically lacking a promoter and an attC recombination site. Gene cassettes can exist as free circular forms, but they lack replicative ability and are commonly integrated into an integron. An integron is defined by an intI gene, an attI recombination site, and a promoter (Pc) [47]. The intI gene encodes a unique tyrosine recombinase specific to sites, with an extra domain that sets it apart from related enzymes. This recombinase facilitates recombination between the integron's attI site and a cassette's attC site, allowing the cassette to be integrated into the integron in a way that permits gene expression controlled by the Pc promoter [49]. Multiple cassettes can be inserted into one integron to form a cassette array, often mistakenly called a “cassette, which can confer resistance to multiple antibiotics. Integrons are categorized into classes based on the sequence of the IntI gene, such as IntI1, IntI2, and IntI3, each having corresponding attI1, attI2, and attI3 sites. Class 1 integron was the first to be identified and is commonly found in clinical isolates resistant to antibiotics [50].

Isa	Tnb	Determinant	Resistance
IS1	Tn9	catA1	Chloramphenicol
IS10	Tn10	tet(B)	Tetracycline
IS26d	Tn4352	aphA1	Kanamycin
	Tn6020	aphA1	Kanamycin
		tet(C)	Tetracycline
		tet(D)	Tetracycline
		catA2	Chloramphenicol
	Tn2003	blaSHV	β -Lactams
		Cfr	Phenicolins/lincosamides/oxazolidinones/pleuromutilins/streptogramin A
IS256e		Cfr	Phenicolins/lincosamides/oxazolidinones/pleuromutilins/streptogramin A
IS50	Tn5	aph(3')-IIa-ble-aph(6)-Ic	Kanamycin, bleomycin, streptomycin
IS903	Tn903	aphA1	Kanamycin
IS1999	Tn1999	blaOXA-48-like	Carbapenems
ISAp11	Tn6330	mcr-1	Colistin
ISEc69		mcr-2	Colistin
ISAs2		blaFOX-5	BLBI
ISAbal4	TnaphA6	aphA6	Kanamycin
ISAbal	Tn2006	blaOXA-23	Carbapenems
		blaOXA-237	Carbapenems
ISAbal25	Tn125	blaNDM	Carbapenems

Table 1.
 Examples of insertion sequences and composite transposons are linked to resistance genes in Gram-negative bacteria [40].

Isa	Tnb	Determinant	Resistance
IS16	Tn1547	vanB1	Vancomycin
IS256		Cfr	Phenicol's/lincosamides/oxazolidinones/pleuromutilins/streptogramin A
	Tn1547	vanB1	Vancomycin
	Tn4001	aacA-aphD	Gentamicin/kanamycin/tobramycin
	Tn5281	aacA-aphD	Gentamicin/kanamycin/tobramycin
	Tn5384	aacA-aphD	Gentamicin/kanamycin/tobramycin
	Tn5384	erm(B)	MLS antibiotics
IS257c		aacD	Kanamycin/neomycin/paromomycin/tobramycin
		aphA-3	Kanamycin/neomycin
		bcrAB	Bacitracin
		Ble	Bleomycin
		dfrK	Trimethoprim
		erm(C)	MLS antibiotics
		fosB5	Fosfomycin
		fusB	Fusidic acid
		ileS2 (<i>mupA</i>)	Mupirocin
		qacC	Antiseptics/disinfectants
		sat4	Streptothricin
		tet(K)	Tetracycline
		tet(L)	Tetracycline
		vat(A)	Streptogramin A
		vga(A)	Streptogramin A/pleuromutilins/lincosamides
		vgb(A)	Streptogramin B

Isa	Tnb	Determinant	Resistance
	Tn924	aacA-aphD	Gentamicin/kanamycin/tobramycin
	Tn4003	dfrA	Trimethoprim
	Tn6072	aacA-aphD	Gentamicin/kanamycin/tobramycin
	Tn6072	Spc	Spectinomycin
IS182	Tn5405	aadE	Spectinomycin

Table 2.
Examples of insertion sequences and composite transposons are linked to resistance genes in Gram-positive bacteria [40].

13.3 Plasmids

Plasmids play a crucial role in transferring mobile genetic elements and acquiring antimicrobial resistance genes in both Gram-negative and Gram-positive bacteria. They vary in size, ranging from less than a kilobase to several megabases. Their ability to replicate independently enables them to be passed on to new host cells, often requiring specific gene systems for efficient vertical inheritance. Some plasmids also facilitate conjugation or mobilization to spread horizontally [51]. These plasmids have a central “backbone” comprising essential maintenance functions, with additional “activities added to benefit the host cell and plasmid in a specific environment, especially in resistance plasmids, which contain resistance genes and mobile elements like insertion sequences, transposons, and integrons. Different backbone structures may have various insertions or resistance regions, with some sharing the same resistance genes and mobile elements [52]. This section will outline the main functions encoded by plasmid backbones and describe key characteristics of known plasmid groups that have significantly contributed to the dissemination of antibiotic resistance within species. Post-replication, plasmids must be distributed to daughter cells during cell division. Small plasmids, present in high numbers, are efficiently inherited by both daughter cells through random segregation [53]. In contrast, larger plasmids present in low copy numbers face the risk of being outcompeted by plasmid-free counterparts unless they contain functional modules for maintenance, such as multimer resolution, partitioning, and post-segregational killing systems. Resolution systems convert plasmid multimers into monomers for independent segregation into daughter cells and usually consist of a site-specific recombinase and a specific DNA site [53]. Some plasmids only have a site recognized by a chromosomally encoded resolvase. Partitioning systems help distribute plasmid copies to daughter cells and typically involve two genes encoding a DNA-binding adaptor protein and a motor protein. Post-segregational killing systems, or plasmid addiction systems, eliminate cells that do not inherit the plasmid due to replication, resolution, or partitioning failures. These systems include toxin-antitoxin systems with a toxic polypeptide and an antitoxin that inhibits toxin expression or activity. Plasmid TA systems rely on a labile antitoxin compared to the toxin, leading to cell death if the plasmid is not inherited by the daughter cell. Additionally, restriction-modification systems on plasmids and other mobile elements can act as post-segregational killing systems [54].

13.4 Transformation

Transformation is the process in which a bacterial cell acquires new genetic material from the environment. There are two models of transformation: (1) Natural transformation, in which bacterial cells use a natural mechanism to absorb DNA from the environment where other bacteria have existed. When the bacteria die and decompose, their DNA is released into the environment. And other bacterial cells can enter them into their system. The way to transfer this foreign DNA is through the bacterial cell membrane and it is combined in the DNA strand. In some cases, if the sequence obtained from the environment is similar to the bacterial sequence, it is combined with the bacterial genome [55]. (2) Transformation in laboratory conditions of foreign DNA absorption occurs through laboratory conditions. Researchers use this method to study genes and create a specific function or remove a system to use. The most common bacterium in this method is *E. coli* bacteria. This bacterium

is easily used in genetic engineering due to its high division speed and has a highly permeable membrane for the entry of foreign DNA. Transformation in laboratory conditions is achieved through electroporation. In this method, bacterial cells are placed in an electric field so that foreign DNA easily passes through the host membrane. Also, another method used in transformation in laboratory conditions is thermal shock, which allows foreign DNA to enter the bacterial genome through the high temperature of the cell membrane. Plasmid is often used to carry out transformation in laboratory conditions and transfer target genes [56]. Plasmids can reproduce in bacterial cells and are used as vectors. Plasmids are widely used for simulating a mechanism and studying it, transfer, and genetic engineering. In many cases, plasmids carry antibiotic-resistance genes and help researchers to identify antibiotic-resistant cells [57].

13.5 Conjugation

Conjugation involves a bacterial cell transmitting genetic information directly to another bacterial cell. The donor cell first replicates part or all of its genetic material and then forms a tube known as a sex pilus between the two cells [58]. The copied DNA can then move through the sex pilus as a single linear strand, and the recipient cell can integrate the DNA into its chromosome or plasmid. While the recipient cell has the potential to receive a copy of every gene from the donor cell, this does not always happen as the fragile sex pilus can break randomly, halting conjugation and preventing the transfer of any remaining genes [59]. Only bacterial cells with a fertility factor can act as donor cells, as possessing a fertility factor gene is a prerequisite for this role, whereas any bacterial cell can serve as a recipient cell. The fertility factor is a gene that enables a bacterial cell to produce a sex pilus [60]. When the fertility factor gene is located on a plasmid, the plasmid is termed an F⁺ plasmid, indicating the presence of a fertility factor. If the fertility factor gene is incorporated into the chromosome, the cell is referred to as an Hfr bacteria [61]. Due to the size of the chromosome, the transfer of a fertility factor within it to a recipient cell is uncommon, with the sex pilus often breaking before this region can be passed on. Fertility factors situated on plasmids are more likely to be transferred to recipient cells compared to those located on chromosomes [62].

13.6 Transduction

Transduction involves a bacterial cell obtaining genetic material from a virus that has infected it. Bacteriophages, a type of virus that infects bacteria, insert their DNA into the bacterial DNA to replicate using the host cell's machinery [63]. This process takes place at a specific location known as the attachment site. During an infection, the bacterial cell replicates the viral DNA and packages it into viral capsids, which are then released from the cell. In some cases, genes near the attachment site may also be mistakenly packaged with the viral DNA, resulting in their transfer to the next bacterial cell infected by the virus [64]. Specialized transduction occurs when only genes adjacent to the attachment site are transferred, while generalized transduction allows for the packaging of any bacterial gene into the new virus. The speed of the infection cycle and viral release varies among viruses, with some replicating rapidly in a lytic stage, while others remain dormant for longer periods in a lysogenic stage. Certain viruses can switch between lytic and lysogenic stages based on the environmental conditions [65].

13.7 Antibiotic-resistant in bacteria

The issue of antibiotic resistance in the treatment of bacterial infections in humans is a significant concern, primarily attributed to the excessive use of antibiotics [66]. The World Health Organization has identified antibiotic resistance as a major challenge of the twenty-first century. Evidence suggests that the administration of antibiotics to animals can lead to the emergence of drug-resistant bacterial infections in both animals and humans [67]. Various measures have been enforced in certain countries like the United States, the United Kingdom, Sweden, and the Netherlands to regulate the utilization of antibiotics in animals [68]. Conversely, in India, a substantial quantity of different antibiotics is routinely employed in farm animals, particularly in the poultry sector. Consequently, there has been a progressive increase in antibiotic resistance among the natural bacteria of animals and their surroundings [69]. Antibiotic-resistant bacteria originating from animals have the potential to transmit to the human population through pathways such as direct contact, consumption of animal meat, and agricultural practices [70]. Resistance genes carried by bacteria can be excreted in feces, facilitating their transfer to other bacterial pathogens. The accumulation of resistant bacteria in animals can result in the dissemination of antibiotic-resistant bacteria to humans [71]. Therefore, concerns have been expressed regarding the use of antibiotics in animal feed, as this practice could contribute to the proliferation of antibiotic-resistant bacteria in humans. The utilization of antibiotics in livestock can lead to contamination of the environment [72]. Consequently, when animals excrete antibiotics that are utilized as manure, these substances can contaminate the environment via agricultural waste. This could facilitate the development and spread of antibiotic-resistant bacteria. In India, antibiotics are extensively employed in food animals, particularly within the poultry industry [73]. However, there are presently no regulatory directives in Iran concerning the administration of antibiotics in animals meant for domestic consumption [73]. Furthermore, antibiotic-resistant bacteria and genes have been detected in poultry litter. For example, *Staphylococcus sciuri*, a prevalent bacterium in poultry, has been shown to possess antibiotic resistance [74].

13.8 Understanding the mechanisms of antibiotic resistance in gram-positive bacteria

The challenge posed by antimicrobial resistance (AMR) is significant, with an estimated annual occurrence of over 670,000 infections caused by AMR pathogens in the European Union (EU), resulting in more than 33,000 fatalities. Additionally, the economic ramifications are substantial, with projections suggesting that the impact on EU healthcare systems may exceed 1 billion euros. The majority of deaths are attributed to a limited number of species, including multi-drug-resistant (MDR) *Escherichia coli*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Acinetobacter* spp., and *Enterococcus* spp. [75]. These bacteria have rapidly developed mechanisms of resistance in response to selective pressures, such as the production of enzymes that deactivate antibiotics, alterations in antibiotic targets, and reduction of intracellular antibiotic levels. Understanding these resistance mechanisms is imperative for healthcare providers when choosing appropriate antibiotic therapies, particularly in cases where laboratory test results are pending [76]. This paper offers an overview of the key resistance mechanisms found in significant AMR pathogens like *S. aureus*, *Enterococcus* spp., *E. coli*, *P. aeruginosa*, *Acinetobacter* spp., and *K. pneumoniae*.

13.9 *S. aureus*

Staphylococcus aureus (*S. aureus*) is considered a significant pathogen largely due to its high levels of resistance to antibiotics [77]. The bacterium has been observed to develop new resistance mechanisms since the early 1940s when penicillin resistance was first detected [78]. As a consequence of this ongoing evolution, *S. aureus* has exhibited resistance to a broad spectrum of antibiotics such as β -lactams, tetracyclines, aminoglycosides, fluoroquinolones, clindamycin, trimethoprim-sulfamethoxazole, vancomycin, daptomycin, and linezolid. In the European Union, Methicillin-resistant *S. aureus* (MRSA), a prevalent strain with antibiotic resistance, was associated with more than 148,000 infections and 7000 fatalities in the year 2023 [78]. The impact of antibiotic resistance in *S. aureus* is overviewed in **Table 3**.

Initially, *S. aureus* resistance to β -lactams arose due to β -lactamases being developed. β -lactam antibiotics target the transpeptidase moiety in penicillin-binding protein (PBP)-2. β -lactams mimic D-Ala4-D-Ala5 and attach to a serine residue in the

Antibiotic	Mechanism of resistance
Penicillins	Penicillinase, production of PBP2a
Cephalosporins generation 1	PBP2a
Cephalosporins generation 2	PBP2a
Cephalosporins generation 3	PBP2a
Cephalosporins generation 4	PBP2a
β -lactamase inhibitor	PBP2a
Carbapenems	PBP2a
Tetracyclines	Ribosomal methylation of binding sites, efflux pumps
Tigecycline	efflux pumps
Macrolides	Ribosomal methylation of binding sites, efflux pumps
Clindamycin	Ribosomal methylation of binding sites, efflux pumps
Fluoroquinolones	Mutations in topoisomerase IV and DNA gyrase, efflux pump
Rifampicin	Mutations in RNA polymerase gene
TMP/SMX	Mutations in DHPS and DHFR
Aminoglycosides	Aminoglycoside degradation enzymes
Aminoglycosides	Electrostatic repulsion through an increase to the cell-surface charge
Vancomycin	VRSA: altered structure of peptidoglycan precursors from D-Ala-D-Ala to D-Ala-D-Lac; VISA: increased production of peptidoglycan, thicker cell wall, decoy D-Ala-D-Ala dipeptides on cell surface
Linezolid	Mutations to the 23S rRNA, altering required modifications to the 23S rRNA, mutations to the 50S ribosomal L3 protein

Table 3.
The impact of antibiotic resistance in *S. aureus* [77].

active site of transpeptidase, forming a stable intermediate known as penicillin-O-serine, which hinders peptidoglycan synthesis. Serine β -lactamases of *S. aureus* form this same intermediate but quickly hydrolyze it, resulting in the release of an inactive product called penicilloic acid. The gene carrying the β -lactamase is situated on a transposon either in a plasmid or integrated into the genome of *S. aureus* [66].

Methicillin-resistant *S. aureus* (MRSA) becomes resistant to methicillin and other β -lactam antibiotics by producing PBP2a, a protein similar to PBP2, but unaffected by β -lactam antibiotics due to its altered active site structure. This resistance extends to most β -lactam antibiotics, except for ceftaroline fosamil. Ceftaroline fosamil can overcome this resistance by binding to an allosteric site on PBP2a, resulting in a conformational change that allows access to the active site serine [79].

Vancomycin resistance in *S. aureus* comes in two forms: vancomycin-resistant *S. aureus* (VRSA) and vancomycin-intermediate *S. aureus* (VISA). Generally, vancomycin resistance is found in MRSA strains. Interestingly, the mechanisms of VISA and VRSA are entirely distinct; hence VISA strains cannot evolve into VRSA, except under rare circumstances with the acquisition of additional resistance mechanisms [80].

Daptomycin, an antibiotic with a unique mechanism, binds to calcium, leading to the formation of cationic complexes that then interact with phosphatidylglycerol (PG) on the cell membrane, eventually causing cell death. Daptomycin's high affinity to PG makes it specific for Gram-positive bacteria, although it is ineffective in pulmonary infections due to issues distinguishing PG in bacteria and lung surfactants. Resistance to daptomycin can occur through mechanisms that alter its interaction with PG [81].

Macrolides and clindamycin hinder protein synthesis by attaching to the 23S rRNA in the 50S ribosomal subunit. Resistance to macrolides may result from ribosomal methylation or efflux pumps expelling the antibiotics from within the cell. Erythromycin ribosomal methylase (*erm*) genes, responsible for resistance to clindamycin, are widespread among *S. aureus* strains and may be expressed constitutively or induced [82].

Clindamycin is a preferred antibiotic for treating MRSA infections in community settings. It is crucial to identify the macrolide, lincosamide, and streptogramin type B (MLSB) resistance phenotype to assess the efficacy of clindamycin therapy. Identifying constitutively resistant MLSB strains (cMLSB) and strains with inducible MLSB (iMLSB) resistance, which may turn resistant during treatment, is vital. The standard D-test can help detect iMLSB resistance in *S. aureus* isolates [83].

Linezolid, an antibiotic belonging to the oxazolidinone class, functions by inhibiting protein synthesis through binding to the 23S rRNA segment in the 50S ribosomal subunit. It is commonly used to combat resistant Gram-positive bacteria like MRSA and VRE. Nonetheless, *S. aureus* has developed resistance mechanisms like mutations in the 23S rRNA and alterations to modifications needed for susceptibility to linezolid [83].

13.10 *Enterococcus* spp.

Enterococcus faecium and *Enterococcus faecalis* are significant pathogens because of their resistance to antibiotics, particularly vancomycin, which can be transmitted to other bacterial species, leading to VRSA [84]. Moreover, they are naturally resilient to various antibiotics. In the European Union in 2023, VRE strains of *E. faecium* and *E. faecalis* caused more than 16,000 infections and 1000 fatalities. *Enterococci* have

penicillin-binding proteins with low affinity, making them inherently resistant to β -lactam antibiotics like penicillin and ampicillin. They are also less responsive to carbapenems and entirely resistant to cephalosporins. The resistance to cephalosporins in enterococci is probably due to the weak affinity of enterococcal penicillin-binding proteins toward antibiotics, specifically PBP5, and other mechanisms such as two-component signal transduction systems [85]. The co-administration of ampicillin and ceftriaxone is effective against *E. faecalis*, similar to ampicillin and gentamicin, which is the standard treatment. Enterococci can mutate their penicillin-binding proteins, resulting in high-level resistance to penicillins, as seen with PBP5 of *E. faecium*. Additionally, enterococci can acquire plasmid-mediated bla genes for the production of β -lactamase, similar to those discovered in *S. aureus* [86]. Despite the known synergy of aminoglycosides and β -lactams, both *E. faecalis* and *E. faecium* are intrinsically resistant to aminoglycosides due to their impermeable cell walls. Nonetheless, when combined with a cell wall synthesis inhibitor like ampicillin, aminoglycosides can penetrate the enterococcal cell and carry out their bactericidal function [87]. Furthermore, *E. faecium* can produce acetyltransferases and phosphotransferases that enzymatically inhibit multiple aminoglycosides, including tobramycin, kanamycin, and amikacin [88]. Hence, gentamicin and streptomycin are the only reliable options against enterococci. Nevertheless, enterococci can acquire mobile genetic elements that confer high-level resistance to these aminoglycosides. Acquired resistance to gentamicin is mediated by enzymes that phosphorylate and acetylate the antibiotic, preventing its binding to the 30S ribosomal subunit [89]. Streptomycin resistance is facilitated by aminoglycoside-modifying enzymes and can also occur through ribosomal mutations, which raise the MIC for streptomycin by eight times compared to the production of aminoglycoside-modifying enzymes. Enterococci have an inherent resistance to TMP-SMX due to their unique ability to absorb folate from their surroundings [90]. They could result in enterococci displaying susceptibility to TMP-SMX in laboratory tests, but this does not necessarily imply that TMP-SMX will be clinically effective against enterococci in living organisms. The resistance of fluoroquinolones in enterococci appears to be mostly acquired, with mechanisms that include mutations in the “quinolone resistance-determining regions” that reduce the enzymes’ affinity for the antibiotic, the production of efflux pumps, and a protective protein encoded by the qnr gene, which is believed to safeguard DNA gyrase by limiting the binding of fluoroquinolones to the enzyme [91]. Additionally, enterococci are inherently resistant to macrolides and clindamycin. Macrolide resistance seems to be influenced by efflux pumps and deactivating enzymes, like an erythromycin methylase encoded by the erm gene [92]. The mechanism of clindamycin resistance varies among species in *E. faecalis*, resistance is due to the presence of the *lsa* gene, which encodes an ATP-energized efflux pump, while in *E. faecium*, resistance appears to be mediated by the *linB* gene, which encodes a nucleotidyltransferase that deactivates clindamycin through adenylation [92]. The development of tetracycline resistance in enterococci involves mechanisms like the production of efflux pumps and ribosomal protection proteins. Overexpression of genes such as tetL and tetM has been linked to resistance to tigecycline as well. The mechanism of daptomycin resistance in *E. faecium* involves the electrostatic repulsion of the daptomycin-calcium complex from the cell membrane, leading to an increase in the membrane’s positive charge. Daptomycin resistance occurs when mutations in the LiaFSR signaling system disrupt the cell membrane and cell wall’s homeostasis, resulting in the redistribution of cardiolipin. This prevents daptomycin from oligomerizing in the septal cell membrane.

The mechanism of vancomycin resistance in enterococci is the same as that found in VRSA and was transferred to *S. aureus* by VRE [93].

14. Mechanisms of antimicrobial resistance in gram-negative bacteria

14.1 *P. aeruginosa*

Pseudomonas aeruginosa, also known as *P. aeruginosa*, is a highly worrisome Gram-negative bacterium that is resistant to multiple drugs, coming in second only to *E. coli* in terms of infections and fatalities caused. In the European Union, it is estimated that multi-drug-resistant *P. aeruginosa* is responsible for more than 72,000 infections and over 4800 deaths every year. More specifically, around 61,892 cases and 4155 deaths are linked to strains resistant to carbapenem, while 1262 cases and 84 deaths are linked to strains resistant to colistin [78]. The mechanisms of antibiotic resistance in *P. aeruginosa* and other significant Gram-negative bacteria are detailed in **Table 4**.

The bacterium's membrane permeability is very low, making it highly resistant to many antibiotics, estimated to be between 12 and 100 times less permeable than *E. coli*'s membrane. Antibiotics that manage to enter the *P. aeruginosa* cell encounter

Antibiotic	Common mechanisms of resistance in Gram-negative bacteria
	<i>P. aeruginosa</i> <i>E. coli</i> <i>K. pneumoniae</i> <i>A. baumannii</i>
Penicillins	AmpC, ESBLs, other b-lactamases
Cephalosporins generation 1	AmpC, ESBLs
Cephalosporins generation 2	AmpC, ESBLs
Cephalosporins generation 3	AmpC, ESBLs
Cephalosporins generation 4	ESBLs
b-lactamase inhibitor	AmpC
Aztreonam	ESBLs
Carbapenems	Class B & D
Tigecycline	Efflux pumps
Macrolides	Efflux pumps
clindamycin	Efflux pumps
Fluoroquinolones	Mutations in topoisomerase IV and DNA gyrase, efflux pump
Rifampicin	Mutations in RNA polymerase gene
TMP/SMX	Efflux Pumps
Aminoglycosides	Aminoglycoside degradation enzymes
Colistin	Reduction of membrane negative charge through addition of N4-aminoarabinose to lipid A

Table 4.
Mechanisms of antibiotic resistance in P. aeruginosa and other important gram-negative bacteria [78].

efflux pumps and antibiotic-inactivating enzymes, reducing their effectiveness. *P. aeruginosa* can develop resistance through mutations that result in deficiencies in porins, decreased permeability, changes to antibiotic targets, or increased expression of efflux pumps and antibiotic-inactivating enzymes [80]. The bacterium also has adaptive mechanisms for antibiotic resistance, like forming biofilms and persister cells. Biofilms protect bacteria from the immune system and environmental stressors, while persister cells can withstand high antibiotic concentrations by halting the synthesis of antibiotic target proteins. *P. aeruginosa* has both natural and acquired mechanisms to resist β -lactam antibiotics [94]. Natural mechanisms include influx pumps and β -lactamases like AmpC and extended-spectrum β -lactamases (ESBLs). Acquired resistance mechanisms involve mutations in porins, overexpression of hydrolyzing enzymes and efflux pumps, and modification of penicillin-binding proteins, reducing susceptibility to various β -lactams. Acquisition of β -lactamases like Class B carbapenemases is linked to high morbidity and mortality from carbapenem-resistant strains of *P. aeruginosa*. AmpC enzymes, being cephalosporinases, can hydrolyze various antibiotics and are inducible, causing quick resistance emergence [95]. Antipseudomonal cephalosporins may not initially induce AmpC expression, but prolonged use can lead to treatment failure due to overproduction of AmpC. ESBLs hydrolyze several antibiotics and are constitutively expressed, meaning certain antibiotics may not lead to therapeutic failure in ESBL-positive strains. OXA-type β -lactamases produced by *P. aeruginosa* can hydrolyze specific antibiotics and are not inhibited by clavulanic acid [96]. Carbapenem resistance in *P. aeruginosa* involves metallo- β -lactamases that can hydrolyze most β -lactams but not aztreonam. Resistance to aminoglycosides in *P. aeruginosa* is due to impermeability. Other resistance mechanisms include overexpression of efflux pumps and changes to the ribosomal target [97]. Resistance to fluoroquinolones in *P. aeruginosa* is related to efflux pumps and mutations in target enzymes encoding topoisomerase IV and DNA gyrase, resulting in high-level resistance. The bacterium is also resistant to tetracyclines, macrolides, TMP/SMX, and rifampicin due to efflux pump production [33]. Efflux pumps are responsible for causing resistance to macrolide and TMP/SMX antibiotics. Mutations in the RNA polymerase gene lead to rifampicin resistance, but this could potentially harm the fitness of *P. aeruginosa* [44]. Colistin is typically used as a final resort to combat multidrug-resistant Gram-negative pathogens, working by disrupting the cell membrane through its interaction with lipid A in the lipopolysaccharide (LPS). By binding to the phosphate groups on lipid A, colistin inserts itself into the membrane, ultimately causing cell lysis by compromising its integrity. *P. aeruginosa* develops resistance to colistin by modifying the negative charge of the cell membrane needed for interaction with the positively charged colistin molecule. This is achieved through the production of N4-aminoarabinose, which binds to lipid A and neutralizes its negative charge [98].

14.2 *E. coli*

Antibiotic-resistant *E. coli* strains were the primary cause of infections and deaths in the EU in 2023, with over 700,000 infections and 8465 fatalities recorded. This had a more significant impact than *S. aureus* and *P. aeruginosa* combined. The third-generation cephalosporin-resistant *E. coli* strains accounted for more than 343,000 infections and 10,000 deaths. Infections caused by colistin-resistant *E. coli* strains resulted in 7156 cases and 500 deaths, while carbapenem-resistant strains led to 2619 infections and 141 deaths [99]. *E. coli* strains can

produce various β -lactamases, including ESBLs, AmpC, and carbapenemases of classes A, B, and D. Initially, SHV-ESBLs and TEM-ESBLs were predominant in *E. coli*, but CTX-M-ESBLs now hold that position [100]. The most commonly found ESBL gene in human *E. coli* isolates is blaCTX-M15, responsible for 88.6% of annual infections caused by third-generation cephalosporin-resistant *E. coli* strains in the EU. The presence of AmpC β -lactamases in *E. coli* has been of interest since the first *E. coli* AmpC β -lactamase was discovered in 1940 [99]. Unlike other Enterobacteriaceae, *E. coli* typically produces low levels of AmpC β -lactamases due to the absence of AmpR. Mutations in the ampC gene or acquisition of a stronger promoter are mechanisms for AmpC hyperproduction [100]. Carbapenem resistance has risen due to ESBL and AmpC-producing strains, necessitating increased use of carbapenems, resulting in the development of carbapenem-resistant strains. This resistance can arise through various mechanisms, including the production of carbapenemases, efflux pumps, and porin mutations. The main carbapenemase variants in Enterobacteriaceae are KPCs, NDM, VIM, IMP, and OXA-48, with NDM and OXA-48 being prevalent in *E. coli* [101]. Resistance to rifampicin, macrolides, fluoroquinolones, and tetracyclines in *E. coli* has similar mechanisms to other Enterobacteriaceae [97]. Mechanisms include mutations in specific genes, target site modifications, and efflux pump production. Aminoglycoside resistance in *E. coli* is mediated through enzymes such as acetyltransferases and phosphotransferases. The resistance to colistin in *E. coli* involves modifications in LPS, with plasmid-mediated resistance being noted through the *mcr-1* gene. This gene causes changes in LPS, reducing the affinity of colistin. MCR-1 has been identified in various countries, while another gene for colistin resistance, *mcr-2*, has been found in Belgium [101].

14.3 *K. pneumoniae*

Klebsiella pneumoniae is inherently immune to penicillin, such as ampicillin, due to the existence of specific first-generation β -lactamases like SHV-1 and TEM-1. It can also resist first- and second-generation cephalosporins through the production of broad spectrum β -lactamases (BSBL). Resistance to third- and fourth-generation cephalosporins can occur by creating ESBLs or AmpC enzymes. Carbapenem resistance is the most critical form of β -lactam resistance in *K. pneumoniae*, mainly due to various carbapenemases like KPCs, NDM, and OXA-type carbapenemases. Apart from carbapenemase production, other mechanisms such as loss of outer membrane proteins and efflux pumps can also contribute significantly to carbapenem resistance, sometimes in conjunction with the overproduction of β -lactamases like AmpC or ESBLs [102]. The modified Hodge test assists in distinguishing between carbapenemase and non-carbapenemase-mediated carbapenem resistance by observing the growth of an indicator strain alongside a suspect strain grown with a carbapenem. It is crucial to note that infections caused by carbapenem-resistant *K. pneumoniae* have a high mortality rate in hospitalized patients, reaching up to 48%. There is an increasing concern regarding the rise of carbapenem resistance in hypervirulent *K. pneumoniae*. In *K. pneumoniae*, resistance to aminoglycosides can occur via different mechanisms, including efflux pumps, aminoglycoside-modifying enzymes like adenylation, acetylation, or phosphorylation, and the production of 16SrRNA methylase, which provides high levels of resistance to various common aminoglycosides [103]. When it comes to fluoroquinolone antibiotics, *K. pneumoniae* can develop

resistance through mutations in specific genes, overproduction of efflux pumps, plasmid-carried qnr genes, and a bifunctional aminoglycoside acetyltransferase. Tetracycline resistance in *K. pneumoniae* is primarily due to the presence of efflux pumps like tetB, while tigecycline resistance is mostly a result of increased acrAB efflux pump activity. However, if sensitive, minocycline demonstrates better antibacterial activity against KPC-producing Enterobacteriaceae, including *Klebsiella* spp., compared to tigecycline. Colistin resistance in *K. pneumoniae* can happen through multiple mechanisms, such as upregulation of the acrAB efflux pump, modifications to lipid A of LPS via certain pathways, or mutations in core genes involved in lipid A maturation. Rifampicin, when used with colistin, can have synergistic effects against KPC-producing *K. pneumoniae*, although resistance may develop due to the production of a specific enzyme. Additionally, *K. pneumoniae* has a notable ability to form biofilms, protecting host defenses and antibiotics, especially in medical devices like catheters [104]. The mechanisms of antimicrobial resistance in *Acinetobacter baumannii* involve the production of various β -lactamases [104]. For instance, all *A. baumannii* strains produce a chromosomally encoded AmpC cephalosporinase, while some may also harbor additional β -lactamases like TEM-1, CTX-M, and SHV variants, with certain strains possessing ESBLs. Sulbactam, an inhibitor of Class A β -lactamases, plays a crucial role in *A. baumannii* infections by inhibiting β -lactamases as well as specific proteins of the bacterium [105]. In cases of *A. baumannii* infections, sulbactam is typically administered in combination with other antibiotics like carbapenems or colistin, and high doses might be necessary to ensure effectiveness. Nevertheless, resistance to sulbactam can still arise in *A. baumannii* strains, particularly those producing certain types of β -lactamases like TEM-1 or metallo- β -lactamases such as VIM or NDM [106]. The most crucial type of β -lactamases produced by *A. baumannii* are the OXA-51-like Class D carbapenemases, which are distinct to this bacterium. OXA-51-like enzymes, which have weak carbapenemase activity, are widespread in *A. baumannii* strains, making them a reliable genetic marker for identification. Other carbapenems like OXA-23, OXA-40, and OXA-58 can also be found in *A. baumannii* strains, either in the chromosome or on plasmids. Tetracycline resistance in *Acinetobacter* spp. is linked to ribosomal protection proteins or efflux pump production, while tigecycline resistance is associated with the AdeABC efflux pump or mutations in specific genes that affect membrane permeability or metabolism [106]. However, AdeABC efflux pumps do not influence susceptibility to minocycline in *A. baumannii* strains, even those resistant to tigecycline [106].

15. Conclusion

Bacteria have become resistant to all antibiotics due to exposure. Understanding resistance mechanisms can help clinicians predict resistance patterns, even for antibiotics not included in the antibiogram, and choose the best antibiotic for the specific pathogen. Although new antibiotics may fight established resistance mechanisms, using them unwisely will lead to the development of resistance against them too.

Conflict of interest

The authors declare that they have no competing interests.

Author details

Mohammad Hashemzadeh^{1,2} and Aram Asareh Zadegan Dezfuli^{1,2*}

1 Infectious and Tropical Diseases Research Center, Health Research Institute, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

2 Faculty of Medicine, Department of Microbiology, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

*Address all correspondence to: aramasareh836@yahoo.com

IntechOpen

© 2024 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Del Duca S, Vassallo A, Mengoni A, Fani R. Microbial genetics and evolution. *Microorganisms*. 2022;**23**(10):1274. DOI: 10.3390/microorganisms10071274
- [2] Xia X. DAMBE6: New tools for microbial genomics, phylogenetics, and molecular evolution. *Journal of Heredity*. 2017;**108**:431-437. DOI: 10.1093/jhered/esx033
- [3] Avram O, Rapoport D, Portugez S, Pupko T. M1CR0B1AL1Z3R—A user-friendly web server for the analysis of large-scale microbial genomics data. *Nucleic Acids Research*. 2019;**47**:W88-W92. DOI: 10.1093/nar/gkz423
- [4] Shapiro JA. Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. *Journal of Molecular Biology*. 1969;**40**:93-105. DOI: 10.1016/0022-2836(69)90298-8
- [5] Torrance EL, Burton C, Diop A, Bobay LM. Evolution of homologous recombination rates across bacteria. *National Academy of Sciences of the United States of America*. 2024;**121**:e2316302121. DOI: 10.1073/pnas.2316302121
- [6] Konstantinidis KT, DeLong EF. Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *The ISME Journal*. 2008;**2**:1052-1065
- [7] Liu Z, Good BH. Dynamics of bacterial recombination in the human gut microbiome. *PLoS Biology*. 2024;**22**(2):e3002472. DOI: 10.1371/journal.pbio.3002472
- [8] Garant D. Natural and human-induced environmental changes and their effects on adaptive potential of wild animal populations. *Evolutionary Applications*. 2020;**13**:1117-1127. DOI: 10.1111/eva.12928
- [9] Noble D, Phillips D. Speciation by physiological selection of environmentally acquired traits. *The Journal of Physiology*. 2024;**602**:2503-2510. DOI: 10.1113/JP285028
- [10] Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecsek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science*. 2020;**367**(6484):eaay5012. DOI: 10.1126/science.aay5012
- [11] Kogay R, Wolf YI, Koonin EV. Defence systems and horizontal gene transfer in bacteria. *Environmental Microbiology*. 2024;**26**:e16630. DOI: 10.1111/1462-2920.16630
- [12] Kehlet-Delgado H, Montoya AP, Jensen KT, Wendlandt CE, Dexheimer C, Roberts M, et al. The evolutionary genomics of adaptation to stress in wild rhizobium bacteria. *National Academy of Sciences of the United States of America*. 2024;**121**:e2311127121. DOI: 10.1073/pnas.2311127121
- [13] Hill MS, Vande Zande P, Wittkopp PJ. Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews Genetics*. 2021;**22**:203-215. DOI: 10.1038/s41576-020-00304-w
- [14] Sanjuan R, Domingo-Calap P. Genetic diversity and evolution of viral populations. *Encyclopedia of Virology*. Elsevier Ltd.; 2021;**1**:53-61. DOI: 10.1016/B978-0-12-809633-8.20958-8

- [15] Spielman SJ, Weaver S, Shank SD, Magalis BR, Li M, Kosakovsky Pond SL. Evolution of viral genomes: Interplay between selection, recombination, and other forces. In: *Evolutionary Genomics: Statistical and Computational Methods*. New York: Springer; 2019. pp. 427-468. DOI: 10.1007/978-1-4939-9074-0_14
- [16] Signor SA, Nuzhdin SV. The evolution of gene expression in cis and trans. *Trends in Genetics*. 2018;**34**:532-544. DOI: 10.1016/j.tig.2018.03.007
- [17] Uller T, Moczek AP, Watson RA, Brakefield PM, Laland KN. Developmental bias and evolution: A regulatory network perspective. *Genetics*. 2018;**209**:949-966. DOI: 10.1534/genetics.118.300995
- [18] Gurney J, Aldakak L, Betts A, Gougat-Barbera C, Poisot T, Kaltz O, et al. Network structure and local adaptation in co-evolving bacteria–phage interactions. *Molecular Ecology*. 2017;**26**:1764-1777. DOI: 10.1111/mec.14008
- [19] Abe K, Takamatsu T, Sato T. Mechanism of bacterial gene rearrangement: SprA-catalyzed precise DNA recombination and its directionality control by SprB ensure the gene rearrangement and stable expression of *spsM* during sporulation in *Bacillus subtilis*. *Nucleic Acids Research*. 2017;**20**(45):6669-6683. DOI: 10.1093/nar/gkx466
- [20] Gomes AL, Johns NI, Yang A, Velez-Cortes F, Smillie CS, Smith MB, et al. Genome and sequence determinants governing the expression of horizontally acquired DNA in bacteria. *The ISME Journal*. 2020;**14**:2347-2357. DOI: 10.1038/s41396-020-0696-1
- [21] Akmayan I, Ozturk AB, Ozbek T. Recombinant proteins production in *Escherichia coli* BL21 for vaccine applications: A cost estimation of potential industrial-scale production scenarios. *Preparative Biochemistry & Biotechnology*. 2024;**6**:1-4. DOI: 10.1080/10826068.2023.2299495
- [22] Lotfi Sheervalilou R, Zarghami N. An update of the recombinant protein expression systems of Cyanovirin-N and challenges of preclinical development. *Bioimpacts*. 2018;**8**:139. DOI: 10.15171/bi.2018.16
- [23] Araújo LS, Silva BB, Santos EN, Bezerra AS, Frota SS, Montenegro AR, et al. Immune response to inactivated bacterial vector carrying the recombinant K39 antigen of *Leishmania infantum* in mice. *Asian Pacific Journal of Tropical Biomedicine*. 2024;**14**:199-206. DOI: 10.4103/apjtb.apjtb_914_23
- [24] Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiology Reviews*. 2014;**38**:865-891. DOI: 10.1111/1574-6976.12067
- [25] Gebrie A. Transposable elements as essential elements in the control of gene expression. *Mobile DNA*. 2023;**14**(1):1-17
- [26] Zhang Y, Gladyshev VN. An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. *Bioinformatics*. 2005;**21**:2580-2589. DOI: 10.1093/bioinformatics/bti400
- [27] Wang X, Xu G, Johnson WA, Qu Y, Yin D, Ramkissoon N, et al. Long sequence insertion via CRISPR/Cas gene-editing with transposase, recombinase, and integrase. *Current Opinion in Biomedical Engineering*. 2023;**28**:100491. DOI: 10.1016/j.cobme.2023.100491
- [28] Wang BB, Xu JZ, Zhang F, Liu S, Liu J, Zhang WG. Review of DNA repair

enzymes in bacteria: With a major focus on AddAB and RecBCD. DNA Repair. 2022;**118**:103389. DOI: 10.1016/j.dnarep.2022.103389

[29] Lima-Noronha MA, Fonseca DL, Oliveira RS, Freitas RR, Park JH, Galhardo RS. Sending out an SOS-the bacterial DNA damage response. Genetics and Molecular Biology. 2022;**45**:e20220107. DOI: 10.1590/1678-4685-GMB-2022-0107

[30] De Ste CM, Vacca I, Kwun MJ, Ralph JD, Bentley SD, Haigh R, et al. Phase-variable methylation and epigenetic regulation by type I restriction–modification systems. FEMS Microbiology Reviews. 2017;**41**:S3–S15. DOI: 10.1093/femsre/fux025

[31] Johnston CD, Cotton SL, Rittling SR, Starr JR, Borisy GG, Dewhirst FE, et al. Systematic evasion of the restriction–modification barrier in bacteria. National Academy of Sciences of the United States of America. 2019;**116**:11454–11459

[32] Rusinov I, Ershova A, Karyagina A, Spirin S, Alexeevski A. Lifespan of restriction-modification systems critically affects avoidance of their recognition sites in host genomes. BMC Genomics. 2015;**16**:1–5. DOI: 10.1186/s12864-015-2288-4

[33] Saki M, Farajzadeh Sheikh A, Seyed-Mohammadi S, Asareh Zadegan Dezfuli A, Shahin M, Tabasi M, et al. Occurrence of plasmid-mediated quinolone resistance genes in *Pseudomonas aeruginosa* strains isolated from clinical specimens in Southwest Iran: A multicentral study. Scientific Reports. 2022;**12**:2296. DOI: 10.1038/s41598-022-06128-4

[34] Porse A, Schønning K, Munck C, Sommer MO. Survival and evolution of a large multidrug resistance

plasmid in new clinical bacterial hosts. Molecular Biology and Evolution. 2016;**33**:2860–2873. DOI: 10.1093/molbev/msw163

[35] Parra B, Cockx B, Lutz VT, Brøndsted L, Smets BF, Dechesne A. Isolation and characterization of novel plasmid-dependent phages infecting bacteria carrying diverse conjugative plasmids. Microbiology Spectrum. 2024;**12**:e02537–e02523. DOI: 10.1128/spectrum.02537-23

[36] Bennett PM. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. British Journal of Pharmacology. 2008;**153**:S347–S357. DOI: 10.1038/sj.bjp.0707607

[37] Ngiam L, Weynberg KD, Guo J. The presence of plasmids in bacterial hosts alters phage isolation and infectivity. ISME Communications. 2022;**2**(1):75. DOI: 10.1038/s43705-022-00158-9

[38] Chevrette MG, Gutiérrez-García K, Selem-Mojica N, Aguilar-Martínez C, Yañez-Olvera A, Ramos-Aboites HE, et al. Evolutionary dynamics of natural product biosynthesis in bacteria. Natural Product Reports. 2020;**37**:566–599. DOI: 10.1039/C9NP00048H

[39] Zhang P, Mao D, Gao H, Zheng L, Chen Z, Gao Y, et al. Colonization of gut microbiota by plasmid-carrying bacteria is facilitated by evolutionary adaptation to antibiotic treatment. The ISME Journal. 2022;**16**:1284–1293. DOI: 10.1038/s41396-021-01171-x

[40] Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile genetic elements associated with antimicrobial resistance. Clinical Microbiology Reviews. 2018;**31**:10–128. DOI: 10.1128/cmr.00088-17

- [41] Babakhani S, Oloomi M. Transposons: The agents of antibiotic resistance in bacteria. *Journal of Basic Microbiology*. 2018;**58**:905-917. DOI: 10.1002/jobm.201800204
- [42] Hegstad K, Mikalsen T, Coque TM, Werner G, Sundsfjord A. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant *Enterococcus faecalis* and *Enterococcus faecium*. *Clinical Microbiology and Infection*. 2010;**16**:541-554. DOI: 10.1111/j.1469-0691.2010.03226.x
- [43] Hashemzadeh M, Dezfuli AA, Nashibi R, Jahangirimehr F, Akbarian ZA. Study of biofilm formation, structure and antibiotic resistance in *Staphylococcus saprophyticus* strains causing urinary tract infection in women in Ahvaz, Iran. *New Microbes and New Infections*. 2021;**39**:100831. DOI: 10.1016/j.nmni.2020.100831
- [44] Gillings MR, Paulsen IT, Tetu SG. Genomics and the evolution of antibiotic resistance. *Annals of the New York Academy of Sciences*. 2017;**1388**:92-107. DOI: 10.1111/nyas.13268
- [45] Harmer CJ, Hall RM. IS 26 and the IS 26 family: Versatile resistance gene movers and genome reorganizers. *Microbiology and Molecular Biology Reviews*. 4 Mar 2024:e00119-e00122
- [46] Ross K, Varani AM, Snesrud E, Huang H, Alvarenga DO, Zhang J, et al. TnCentral: A prokaryotic transposable element database and web portal for transposon analysis. *MBio*. 2021;**12**:10-128. DOI: 10.1128/mbio.02060-21
- [47] Shahi F, Hamidi H, Khoshnood S, Mehdipour G, Dezfouli AA, Sheikh AF. Virulence determinants and biofilm formation in clinical isolates of *Enterococcus*: A cross-sectional study. *Journal of Acute Disease*. 2020;**9**:27-32. DOI: 10.4103/2221-6189.276079
- [48] Farajzadeh Sheikh A, Asareh Zadegan Dezfuli A, Navidifar T, Fard SS, Dehdashtian M. Association between biofilm formation, structure and antibiotic resistance in *Staphylococcus epidermidis* isolated from neonatal septicemia in Southwest Iran. *Infection and Drug Resistance*. 2019;**12**:1771-1782. DOI: 10.2147/IDR.S204432
- [49] Fonseca ÉL, Vicente AC. Integron functionality and genome innovation: An update on the subtle and smart strategy of integrase and gene cassette expression regulation. *Microorganisms*. 2022;**20**(10):224. DOI: 10.3390/microorganisms10020224
- [50] Moura A, Soares M, Pereira C, Leitão N, Henriques I, Correia A. INTEGRALL: A database and search engine for integrons, integrases and gene cassettes. *Bioinformatics*. 2009;**25**(8):1096-1098. DOI: 10.1093/bioinformatics/btp105
- [51] Rhee MS, Kim JW, Qian Y, Ingram LO, Shanmugam KT. Development of plasmid vector and electroporation condition for gene transfer in sporogenic lactic acid bacterium, *Bacillus coagulans*. *Plasmid*. 2007;**58**:13-22. DOI: 10.1016/j.plasmid.2006.11.006
- [52] Loyola Irizarry HG, Brito IL. Characterizing conjugative plasmids from an antibiotic-resistant dataset for use as broad-host delivery vectors. *Frontiers in Microbiology*. 2023;**14**:1199640. DOI: 10.3389/fmicb.2023.1199640
- [53] Sivasankar C, Hewawaduge C, Lee JH. Novel pro-and eukaryotic expression plasmid expressing omicron antigens delivered via *Salmonella* elicited MHC class I and II based protective

- immunity. Journal of Controlled Release. 2023;357:404-416. DOI: 10.1016/j.jconrel.2023.04.015
- [54] Yang Q, Zhu Y, Schwarz S, Wang L, Liu W, Yang W, et al. Integrative and conjugative elements in *Streptococci* can act as vectors for plasmids and translocatable units integrated via IS1216E. International Journal of Antimicrobial Agents. 2023;61(5):106793
- [55] Christi K, Hudson J, Egan S. Current approaches to genetic modification of marine bacteria and considerations for improved transformation efficiency. Microbiological Research. 2024;284:127729. DOI: 10.1016/j.micres.2024.127729
- [56] Mårli MT, Oppegård O, Porcellato D, Straume D, Kjos M. Genetic modification of *Streptococcus dysgalactiae* by natural transformation. Msphere. 2024;9:e00214-e00224. DOI: 10.1128/msphere.00214-24
- [57] Lu J, Liu Y, Song M, Xi Y, Yang H, Liu W, et al. The CsPbs2-interacting protein oxalate decarboxylase CsOxdC3 modulates morphosporogenesis, virulence, and fungicide resistance in *Colletotrichum siamense*. Microbiological Research. 2024;284:127732. DOI: 10.1016/j.micres.2024.127732
- [58] Piscon B, Pia Esposito E, Fichtman B, Samburski G, Efremushkin L, Amselem S, et al. The effect of outer space and other environmental cues on bacterial conjugation. Microbiology Spectrum. 2023;11:e03688-e03622. DOI: 10.1128/spectrum.03688-22
- [59] Luque A, Paytubi S, Sánchez-Montejo J, Gibert M, Balsalobre C, Madrid C. Crosstalk between bacterial conjugation and motility is mediated by plasmid-borne regulators. Environmental Microbiology Reports. 2019;11:708-717. DOI: 10.1111/1758-2229.12784
- [60] Graf FE, Palm M, Warringer J, Farewell A. Inhibiting conjugation as a tool in the fight against antibiotic resistance. Drug Development Research. 2019;80:19-23. DOI: 10.1002/ddr.21457
- [61] Wierzbka AJ, Maximova K, Wincenciuk A, Równicki M, Wojciechowska M, Nexø E, et al. Does a conjugation site affect transport of vitamin B12-peptide nucleic acid conjugates into bacterial cells? Chemistry-A European Journal. 2018;24:18772-18778. DOI: 10.1002/chem.201804304
- [62] Ferreira VA, Saraiva MM, Campos IC, da Silva MP, Benevides VP, Almeida AM, et al. In vitro conjugation of Inc B/O-plasmid: Minimum inhibitory concentration of β -lactams increases 16-fold in *Salmonella enterica* compared with *Escherichia coli*. Microbial Pathogenesis. 2024;8:106788. DOI: 10.1016/j.micpath.2024.106788
- [63] Keller CM, Kendra CG, Bruna RE, Craft D, Pontes MH. Genetic modification of *Sodalis* species by DNA transduction. Msphere. 2021;24(6):10-128. DOI: 10.1128/msphere.01331-20
- [64] Marinus MG, Poteete AR. High efficiency generalized transduction in *Escherichia coli* O157: H7. F1000Research. 2013;2:7. DOI: 10.12688/f1000research.2-7v1
- [65] Tiwari S, Jamal SB, Hassan SS, Carvalho PV, Almeida S, Barh D, et al. Two-component signal transduction systems of pathogenic bacteria as targets for antimicrobial therapy: An overview. Frontiers in Microbiology. 2017;8:1878. DOI: 10.3389/fmicb.2017.01878

- [66] Abbasi Montazeri E, Seyed-Mohammadi S, Asarehzadegan Dezfuli A, Khosravi AD, Dastoorpoor M, Roointan M, et al. Investigation of SCC mec types I–IV in clinical isolates of methicillin-resistant coagulase-negative *Staphylococci* in Ahvaz, Southwest Iran. *Bioscience Reports*. 2020;**40**:BSR20200847. DOI: 10.1042/BSR20200847
- [67] Noman EA, Mohamed RM, Al-Gheethi AA, Al-Shaibani MM, Al-Wrafiy FA, Al-Maqtari QA, et al. Antibiotics and antibiotic-resistant bacteria in greywater: Challenges of the current treatment situation and predictions of future scenario. *Environmental Research*. 2022;**212**:113380. DOI: 10.1016/j.envres.2022.113380
- [68] Nelson RE, Hatfield KM, Wolford H, Samore MH, Scott RD, Reddy SC, et al. National estimates of healthcare costs associated with multidrug-resistant bacterial infections among hospitalized patients in the United States. *Clinical Infectious Diseases*. 2021;**72**:S17–S26. DOI: 10.1093/cid/ciaa1581
- [69] Datta S, Nag S, Roy DN. Biofilm-producing antibiotic-resistant bacteria in Indian patients: A comprehensive review. *Current Medical Research and Opinion*. 2024;**40**(3):403–422. DOI: 10.1080/03007995.2024.2305241
- [70] Lenart-Boroń A, Stankiewicz K, Czernecka N, Ratajewicz A, Bulanda K, Heliasz M, et al. Wounds of companion animals as a habitat of antibiotic-resistant bacteria that are potentially harmful to humans—phenotypic, proteomic and molecular detection. *International Journal of Molecular Sciences*. 2024;**25**:3121. DOI: 10.3390/ijms25063121
- [71] Khosravi AD, Tae S, Dezfali AA, Meghdadi H, Shafie F. Investigation of the prevalence of genes conferring resistance to carbapenems in *Pseudomonas aeruginosa* isolates from burn patients. *Infection and Drug Resistance*. 2019;**12**:1153–1159. DOI: 10.2147/IDR.S197752
- [72] Skandalis N, Maeusli M, Papafotis D, Miller S, Lee B, Theologidis I, et al. Environmental spread of antibiotic resistance. *Antibiotics*. 2021;**10**:640. DOI: 10.3390/antibiotics10060640
- [73] Gholami-Ahangaran M, Moravvej AH, Safizadeh Z, Nogoorani VS, Zokaei M, Ghasemian SO. The evaluation of ESBL genes and antibiotic resistance rate in *Escherichia coli* strains isolated from meat and intestinal contents of Turkey in Isfahan, Iran. *Iranian Journal of Veterinary Research*. 2021;**22**:318. DOI: 10.22099/ijvr.2021.39493.5737
- [74] Gómez-Sanz E, Haro-Moreno JM, Jensen SO, Roda-García JJ, López-Pérez M. The resistome and mobilome of multidrug-resistant *Staphylococcus sciuri* C2865 unveil a transferable trimethoprim resistance gene, designated dfrE, spread unnoticed. *Msystems*. 2021;**6**:10–128. DOI: 10.1128/msystems.00511-21
- [75] Huang J, Zhong XF, Gao YZ. New antibiotic against multi-drug resistant bacteria. *The Innovation Life*. 2024;**2**:100057. DOI: 10.59717/j.xinn-life.2024.100057
- [76] Kumar A, Kumar A, Prasad KS. Antibiotic-resistant bacteria in municipal sewage water joining river Ganga, at Prayagraj (India). *Gene Reports*. 2021;**23**:101175. DOI: 10.1016/j.genrep.2021.101175
- [77] Ye J, Chen X. Current promising strategies against antibiotic-resistant

bacterial infections. Antibiotics. 2022;**12**:67. DOI: 10.3390/antibiotics12010067

[78] Kakoullis L, Papachristodoulou E, Chra P, Panos G. Mechanisms of antibiotic resistance in important gram-positive and gram-negative pathogens and novel antibiotic solutions. Antibiotics. 2021;**10**:415. DOI: 10.3390/antibiotics10040415

[79] Dargahi Z, Hamad AA, Sheikh AF, Ahmad Khosravi N, Samei Fard S, Motahar M, et al. The biofilm formation and antibiotic resistance of bacterial profile from endotracheal tube of patients admitted to intensive care unit in Southwest of Iran. PLoS One. 2022;**17**:e0277329. DOI: 10.1371/journal.pone.0277329

[80] Ghapanvari P, Taheri M, Jalilian FA, Dehbashi S, Dezfuli AA, Arabestani MR. The effect of nisin on the biofilm production, antimicrobial susceptibility and biofilm formation of *Staphylococcus aureus* and *Pseudomonas aeruginosa*. European Journal of Medical Research. 2022;**27**:173. DOI: 10.1186/s40001-022-00804-x

[81] Reed JM, Gardner SG, Mishra NN, Bayer AS, Somerville GA. Metabolic interventions for the prevention and treatment of daptomycin non-susceptibility in *Staphylococcus aureus*. Journal of Antimicrobial Chemotherapy. 2019;**74**:2274-2283. DOI: 10.1093/jac/dkz194

[82] El Mammery A, Ramírez de Arellano E, Cañada-García JE, Cercenado E, Villar-Gómara L, Casquero-García V, et al. An increase in erythromycin resistance in methicillin-susceptible *Staphylococcus aureus* from blood correlates with the use of macrolide/lincosamide/streptogramin

antibiotics. EARS-net Spain (2004-2020). Frontiers in Microbiology. 2023;**14**:1220286. DOI: 10.3389/fmicb.2023.1220286

[83] Ahmadkhosravi N, Khosravi AD, Dezfuli AZ, Hashemzadeh M, Saki M, Mehr FJ, et al. Study of aerobic and anaerobic bacterial profile of nosocomial infections and their antibiotic resistance in a referral center, Southwest Iran: A three year cross-sectional study. PLoS One. 2021;**16**:e0259512. DOI: 10.1371/journal.pone.0259512

[84] Kraszewska Z, Skowron K, Kwiecińska-Piróg J, Grudlewska-Buda K, Przekwas J, Wiktorczyk-Kapischke N, et al. Antibiotic resistance of *Enterococcus* spp. isolated from the urine of patients hospitalized in the University Hospital in North-Central Poland, 2016-2021. Antibiotics. 2022;**11**:1749. DOI: 10.3390/antibiotics11121749

[85] Boccella M, Santella B, Pagliano P, De Filippis A, Casolaro V, Galdiero M, et al. Prevalence and antimicrobial resistance of *Enterococcus* species: A retrospective cohort study in Italy. Antibiotics. 2021;**10**:1552. DOI: 10.3390/antibiotics10121552

[86] Martins E, Novais C, Freitas AR, Dias AR, Ribeiro TG, Antunes P, et al. Filling the map for antimicrobial resistance in sub-Saharan Africa: Ampicillin-resistant *Enterococcus* from non-clinical sources in Angola. Journal of Antimicrobial Chemotherapy. 2015;**70**:2914-2916. DOI: 10.1093/jac/dkv172

[87] Mwikuma G, Kainga H, Kallu SA, Nakajima C, Suzuki Y, Hang'ombe BM. Determination of the prevalence and antimicrobial resistance of *Enterococcus faecalis* and enterococcus faecium associated with poultry in four districts in Zambia. Antibiotics. 2023;**12**:657. DOI: 10.3390/antibiotics12040657

- [88] Iancu AV, Arbune M, Zaharia EA, Tutunaru D, Maftai NM, Peptine LD, et al. Prevalence and antibiotic resistance of *Enterococcus* spp.: A retrospective study in hospitals of Southeast Romania. *Applied Sciences*. 2023;**13**:3866. DOI: 10.3390/app13063866
- [89] Georges M, Odoyo E, Matano D, Tiria F, Kyany AC, Mbawika D, et al. Determination of *Enterococcus faecalis* and *Enterococcus faecium* antimicrobial resistance and virulence factors and their association with clinical and demographic factors in Kenya. *Journal of Pathogens*. 2022;**2022**:3129439. DOI: 10.1155/2022/3129439
- [90] Sirichoat A, Flórez AB, Vázquez L, Buppasiri P, Panya M, Lulitanond V, et al. Antibiotic resistance-susceptibility profiles of *Enterococcus faecalis* and *Streptococcus* spp. from the human vagina, and genome analysis of the genetic basis of intrinsic and acquired resistances. *Frontiers in Microbiology*. 2020;**11**:1438. DOI: 10.3389/fmicb.2020.01438
- [91] Sinel C, Cacaci M, Meignen P, Guérin F, Davies BW, Sanguinetti M, et al. Subinhibitory concentrations of ciprofloxacin enhance antimicrobial resistance and pathogenicity of *Enterococcus faecium*. *Antimicrobial Agents and Chemotherapy*. 2017;**61**:10-128. DOI: 10.1128/AAC.02763-16
- [92] Barbosa-Ribeiro M, Gomes BP, Arruda-Vasconcelos R, de Albuquerque MI, Costa MJ, Sette-de-Souza PH. Antibiotic resistance profile of clinical strains of *Enterococci* from secondary/persistent endodontic infections: What do we know? A systematic review of clinical studies. *Journal of Endodontics*. 2024;**50**(3):299-309. DOI: 10.1016/j.joen.2023.12.007
- [93] Munita JM, Mishra NN, Alvarez D, Tran TT, Diaz L, Panesso D, et al. Failure of high-dose daptomycin for bacteremia caused by daptomycin-susceptible *Enterococcus faecium* harboring LiaSR substitutions. *Clinical Infectious Diseases*. 2014;**59**:1277-1280. DOI: 10.1093/cid/ciu642
- [94] Moogahi S, Beni FT, Hashemzadeh M, Dezfuli AA. Molecular identification and biofilm formation of aerobic and anaerobic coinfection bacterial isolated from cystic fibrosis patients in Southwest Iran from 2014 to 2022. *Molecular Biology Reports*. 2023;**50**:508225-508235. DOI: 10.1007/s11033-023-08724-7
- [95] Kazmierczak KM, de Jonge BL, Stone GG, Sahm DF. Longitudinal analysis of ESBL and carbapenemase carriage among Enterobacterales and *Pseudomonas aeruginosa* isolates collected in Europe as part of the international network for optimal resistance monitoring (INFORM) global surveillance programme, 2013-17. *Journal of Antimicrobial Chemotherapy*. 2020;**75**:1165-1173. DOI: 10.1093/jac/dkz571
- [96] Skoglund E, Abodakpi H, Rios R, Diaz L, De La Cadena E, Dinh AQ, et al. *In vivo* resistance to Ceftolozane/Tazobactam in *Pseudomonas aeruginosa* arising by AmpC- and non-AmpC-mediated pathways. *Case Reports in Infectious Diseases*. 2018;**2018**:9095203. DOI: 10.1155/2018/9095203
- [97] Kim YJ, Huh HJ, Sung H. Challenges of carbapenem-resistant *Pseudomonas aeruginosa* in infection control and antibiotic management. *Annals of Laboratory Medicine*. 2024;**44**(1):1-2. DOI: 10.3343/alm.2024.44.1.1
- [98] Wang Y, Kong J, Zhang X, Liu Y, Huang Z, Yuan L, et al. Plumbagin resurrect colistin susceptible against colistin-resistant *Pseudomonas aeruginosa* *in vitro* and *in vivo*. *Frontiers in*

Microbiology. 2022;**13**:1020652. DOI: 10.3389/fmicb.2022.1020652

[99] Okhuysen PC, DuPont HL. Enteraggative *Escherichia coli* (EAEC): A cause of acute and persistent diarrhea of worldwide importance. The Journal of Infectious Diseases. 2010;**204**:503-505. DOI: 10.1086/654895

[100] Colmenarejo C, Hernández-García M, Muñoz-Rodríguez JR, Huertas N, Navarro FJ, Mateo AB, et al. Prevalence and risks factors associated with ESBL-producing faecal carriage in a single long-term-care facility in Spain: Emergence of CTX-M-24-and CTX-M-27-producing *Escherichia coli* ST131-H 30R. Journal of Antimicrobial Chemotherapy. 2020;**75**:2480-2484. DOI: 10.1093/jac/dkaa219

[101] Ba X, Guo Y, Moran RA, Doughty EL, Liu B, Yao L, et al. Global emergence of a hypervirulent carbapenem-resistant *Escherichia coli* ST410 clone. Nature Communications. 2024;**15**:494. DOI: 10.1038/s41467-023-43854-3

[102] Ejaz H, Younas S, Qamar MU, Junaid K, Abdalla AE, Abosalif KO, et al. Molecular epidemiology of extensively drug-resistant mcr encoded colistin-resistant bacterial strains co-expressing multifarious β -lactamases. Antibiotics. 2021;**10**:467. DOI: 10.3390/antibiotics10040467

[103] Harbaoui S, Ferjani S, Mohamed Salah A, Guzmán-Puche J, Causse M, Elías-López C, et al. Genetic background of aminoglycoside-modifying enzymes in various genetic lineages of clinical aminoglycosides-resistant *E. coli* and *K. pneumoniae* isolates in Tunisia. Journal of Applied Microbiology. 2024;**2**:lxae164. DOI: 10.1093/jambio/lxae164

[104] Sonnevend A, Ghazawi A, Darwish D, AlDeesi Z, Kadhum AF,

Pál T. Characterization of KPC-type carbapenemase-producing *Klebsiella pneumoniae* strains isolated in the Arabian peninsula. Journal of Antimicrobial Chemotherapy. 2015;**70**:1592-1593. DOI: 10.1093/jac/dku576

[105] Castanheira M, Kimbrough JH, DeVries S, Mendes RE, Sader HS. Trends of β -lactamase occurrence among *Escherichia coli* and *Klebsiella pneumoniae* in United States hospitals during 5 years and activity of antimicrobial agents against isolates stratified by β -lactamase type. Open Forum Infectious Diseases. 2023;**10**(2):ofad038. DOI: 10.1093/ofid/ofad038

[106] Hajikhani B, Sameni F, Ghazanfari K, Abdolali B, Yazdanparast A, Dezfuli AA, et al. Prevalence of blaNDM-producing *Acinetobacter baumannii* strains isolated from clinical samples around the world; a systematic review. Gene Reports. 2023;**30**:101728. DOI: 10.1016/j.genrep.2022.101728

Section 3

Applications and Case Studies in Human Disease

Chapter 8

Asymmetry of Brain and Behavior of People with Williams Syndrome in Verbal and Nonverbal Processing

*Ching-Fen Hsu, Qian Jiang, Shi-Yu Rao, Xiu-Zhi Fu
and Xin Zheng*

Abstract

People with Williams syndrome (WS) are a population with genetic deficits on chromosome 7q11.23. This deficit is rare and was first identified in 1961. Due to this deficit, the cognitive and linguistic profiles of people with WS are characteristic. The cognitive style of people with WS in information processing is local focusing and global ignorance. People with WS have rich lexical, semantic, and grammatical knowledge, such as conditionals, but show atypical sentence processing and compromised brainwaves on semantic integration. This peculiarity has driven my laboratory to conduct a series of studies to determine the root cause of atypical processing in people with WS in the verbal and nonverbal domains. In this chapter, the findings of our laboratory on verbal and nonverbal domains in people with WS regarding asymmetry of the brain and behavior are introduced. All findings may contribute to the rehabilitation of this population in the near future.

Keywords: Williams syndrome, rehabilitation, language, cognition, asymmetry of brain and behavior

1. Introduction

Language is a communication tool that involves at least one sender and receiver in conveying messages. Messages conveyed successfully by a sender and receiver succeed in communication. Successful communication includes speech, language, and extralinguistic components [1]. Any breakdown in each component results in unsuccessful communication; hence, messages cannot be completely sent to the receivers. In speech components, communication involves articulation, resonance, aspiration, and phonation. While articulating, message senders should clearly convey the manner, place, and voicing of articulation. To produce clear articulation, speakers should make appropriate aspirations with the diaphragm rising up during exhalation and lowering down during inhalation. Upon exhalation, the airstream resonates in the oral or nasal cavity based on the produced sounds. The velum is raised for oral sounds

and lowered for nasal sounds. Phonation is also important for the vibration of sounds in the pharynx and larynx. All the components make speech clear and possible in communication. Communication in the language component includes form, content, and use. Language form refers to the syntactic structure, morphology, and phonology. Language content refers to the semantics of a language. Language use refers to the pragmatics of language or appropriate use in contexts. Only the right context can convey messages meaningfully. In the extralinguistic components, communication includes paralinguistics, nonlinguistics, and metalinguistics. Paralinguistics is required in communication, such as stress, pitch, intonation, prosody, and suprasegmental features. These features are essential in distinct languages; for instance, stress is important in English, and pitch is important in Chinese. Intonation is key in expressing speech acts, such as directive, imperative, or interrogative acts. Prosody is the combination of stress, pitch, and intonation. Suprasegmental features, that is, the features superimposed on segments, are essential in producing correct pronunciation, such as nasalization. Nonlinguistics in communication includes proxemics and kinesics. Proxemics refers to the distance between speakers and listeners. To ensure effective communication, the distance should be comfortable on both sides. Kinesics refers to nonverbal body language, gestures, and facial expressions used to signal messages. Metalinguistics in communication refers to the involvement of thinking and planning of meaningful messages. In addition to the components of speech, language, and extralinguistics, social interaction forms the basis of human communication. From infancy, babies develop an attachment to their caregivers, especially their mothers. This attachment involves infant-directed speech, eye gaze, cuddling, joint attention, and responses to babies. Through attachment, social relationships develop, and an early mode of social interaction develops.

Social ability develops from interactions with the environment. According to neuroconstructivism [2, 3], organisms grow by actively interacting with their environment from embryos to adulthood. Westermann et al. [4] provide robust evidence to support neuroconstructivism. Genes are expressed to bear some characteristics within their surroundings, and encephalization facilitates neuronal signal transmission *via* synapses. Encephalization compensates for cognitive abilities, such as neural plasticity, and embodiment shows motor fitness with contextual complexity. Ensocialment builds up social relationships through the development of attachment. Neuroconstructivism highlights dynamic interactions in the development of organisms and their environments. However, the extent to which gene deficits influence the development of cognitive and language abilities remains unclear [5].

2. Language and cognitive profiles of people with Williams syndrome

To determine the answer to this question, people with neurodevelopmental disorders with gene deficits were studied in our laboratory, including Williams syndrome (WS), Down syndrome (DS), and Prader-Willi syndrome (PWS). Since our laboratory has been conducting research on people with WS the longest compared to other syndromes, in this chapter, people with WS will be focused on and introduced.

Since 1961, Williams syndrome has been identified by cardiologist Dr. Williams [6]. People with WS are characterized by unique genotypes and phenotypes. The genotypes of people with WS are rooted in chromosome 7q11.23 with an average of 15–22 missing genes, including the elastin gene and LIMK1 gene [5, 7]. Due to genetic deficits, people with WS have supravalvular aortic stenosis, hypercalcemia, wide

mouths, large ears, and epicanthic folds. Williams syndrome is a rare disorder with an etiology of 1 in 20,000 to 25,000 live births [8]. People with WS are born with an atypical amygdala and misaligned central fissures between the two hemispheres from typically developing (TD) controls. A smaller anterior frontal region but a larger posterior neocerebellum region than TD is also observed in people with WS. Moreover, people with WS have smaller lenticular nuclei than those with TD. The phenotypes of people with WS include mental disability with an average IQ of 55 and uneven cognitive profiles of impaired visuospatial ability, but relatively good language ability. However, the face processing of people with WS is in the normal range. Based on Bellugi et al. [9], people with WS recognize model faces from several choices in different shades and orientations; however, they cannot identify line angles from various line orientations. This difference in differentiating faces from geometric lines suggests that faces are unique social cues for people with WS. Another study on the social interaction of people with WS showed their attraction to the faces of the experimenters while the experimenters approached them with toys making sounds in their hands. Hence, fluent language, advantageous face processing, and hypersocial tendencies in people with WS lead to the hypothesis of a social module [10].

Recently, our laboratory of language pathology and developmental neurosciences has attempted to rehabilitate people with neurodevelopmental disabilities (people with WS included) by conducting a series of studies on cognition and language. Our findings challenge previous reports of completely impaired visuospatial ability and good language ability in people with WS. In fact, people with WS develop visuospatial ability at the mental age level [11–13] but mixed levels of language ability [14–20]. The face processing of people with WS reaches their chronological age (CA) [21]. Most interestingly, people with WS could perform normal-like behaviors with atypical neurological brainwaves in the verbal [18, 22] and nonverbal domains [21]. These findings of asymmetry between the brain and behavior make the cognitive and language profiles of people with WS different. Further studies using animations and virtual reality make the rehabilitation of people with WS possible in the future [23, 25, 26]. This chapter highlights the implications of brain and behavioral asymmetry on the rehabilitation potential of people with WS to lend support to neuroconstructivism [2, 3, 27, 28].

3. Brain and behavioral asymmetry in verbal domain

The distinct patterns of social behaviors of people with WS might be rooted in the cognitive style of ignorance of the global configuration, but focusing on local information. In a copy drawing study of a two-story house [9], people with WS drew pieces of the house scattered in different places without a complete picture of the house. In another copy drawing study of a bicycle, people with WS drew pieces of the bicycle without forming an entire picture. A study using the Navon test on people with WS revealed the same cognitive profile. In a test of copying a large letter D with small y elements, people with WS drew small letters in two horizontal or vertical lines, without configuration D. This asymmetry in local and global information processing in the nonverbal domain of people with WS is compatible with their rich lexical semantics and deficient sentential processing in the verbal domain [15–17].

To investigate the root cause of deficient sentential processing in people with WS, two hypotheses were generated. The first hypothesis is that people with WS truly have impaired lexical semantics. In other words, people with WS have difficulty

interpreting lexical words in context. The second hypothesis is that people with WS have impaired integration of lexical words while processing sentences. To examine the first hypothesis, we reviewed previous studies. Studies have reported that people with WS have advantageous knowledge of lexical semantics. In a word-generation study, people with WS produced significantly more lexical words in a category, for instance, animals, than TD controls. People with WS produced significantly more low-frequency animals and a greater total number of animals than TD controls. In a study of comprehension of homonyms [28], people with WS picked up both the primary meanings and secondary meanings of homonyms, equally frequent in collocation with the targets, such as bank in its primary meaning of money and secondary meaning of river. In the same study, TD chose the primary meanings of homonyms but not secondary meanings. In Tyler et al. [29], people with WS primed categorical and functionally related word pairs as TD controls. These findings suggest that people with WS have rich lexical semantics, given their intellectual disabilities. Hence, it seems that the first hypothesis was challenged, as people with WS develop at or even exceed normal levels.

To test the second hypothesis, Hsu [15] conducted a study using the classic homonym paradigm across nouns and verbs in people with WS. A leading context was presented before the homonym, followed by a picture corresponding to one of the meanings of the homonym. Immediately after the presentation of the homonym, people with WS had to decide whether the picture was compatible with the leading context. The correct response to each homonym was figurative but not literal. For instance, a target homonym *po1-leng3-shui3* (泼冷水 in Chinese original text) has the literal meaning of ‘dampen cold water’ and the figurative meaning of ‘show no care or be apathy.’ A minor goal was to examine whether the integration of lexical words differed across nouns and verbs. People with WS showed significantly higher yes responses to the literal meaning of verbs compared to the figurative meanings of the same verbs, whereas the reverse pattern was observed in TD controls. People with WS showed a pattern similar to nouns in TD. The findings not only showed degraded integration of lexical words into sentences but also confirmed the different processing of nouns and verbs in people with WS.

Another study on causal inference by Hsu [17] showed delayed backward inference and deviant forward inference in people with WS. In the backward causal inference study, people with WS deviated in causal processing from consequence to cause. In the forward causal inference study, people with WS had delayed causal processing from cause to consequence. The MA controls showed similar performance as the CA controls in backward causal inference but differed from the CA controls in forward causal inference. These findings indicate that people with WS are deficient in higher-level cognitive processing of causal inference. The findings in the study of Hsu and Tzeng [16] uncovered impaired integration of propositions embedded in sentences. These studies suggest that people with WS are deficient in sentence processing. Hsu et al. [22], who conducted a study using the false memory paradigm, reported that people with WS made similar amounts of false alarms to lures as TD controls. The same pattern was observed in Hsu [18] using functional neuroimaging, showing a similar pattern of processing word pairs of causal relations and functional relations. Hence, the second hypothesis confirms that people with WS develop delayed and deviant integration of lexical words into sentences. This might be the root cause of impairment in the sentential processing of people with WS.

Considering the neurological performance in verbal studies, asymmetry of the brain and behavior in people with WS has been observed. In a study that used a false

memory paradigm to examine conceptual integration, while TD controls reflected brainwaves to semantically related lures as presented words, people with WS showed brainwaves to lures as new words [22]. This finding suggests that people with WS processed lures with atypical underlying neurological mechanisms from TD controls. A study on the processing of distinct semantic relationships of causal, categorical, associative, and functional word pairs in people with WS revealed different patterns from TD at the neurological level [18]. By using the subtraction method of TD and people with WS, in processing categorical word pairs, while TD controls processed this type of word pair with the right parietal precuneus, people with WS did not show activation in this brain region. In processing associative word pairs, while TD controls processed words with brain activation in the region of the left frontal subgyral areas, people with WS did not show this activation. In processing causal word pairs, while TD processed words in the left superior frontal gyrus, people with WS did not show this pattern. In processing functional word pairs, while TD processed words in the right insula, people with WS did not show this activation. In summary, people with WS have distinct underlying neurological mechanisms for processing lexical words with distinct semantic relationships. Using distinct neurological mechanisms, however, people with WS show normal-like behavioral performance. This sheds light on the rehabilitation potential of people with WS.

4. Brain and behavioral asymmetry in nonverbal domain

In the nonverbal domain, the asymmetry of the brain and behavior is also valid. People with WS showed problems recognizing faces with emotions because of their inability to integrate pieces of faces related to emotions [30]. Hsu and Chen [21] conducted a study to examine the processing of changes in figures and configures of faces after viewing model faces. Participants were first presented with a model face, followed by a target face with changes in figures or configurations. Participants were asked to judge whether the displayed target was the same or different from the model face. While decisions were made, the participants' brainwaves were recorded using event-related potentials. The results revealed distinct patterns of brain waves in the left and right hemispheres. In the left hemisphere, TD controls differentiated the three types of faces distinctly, whereas people with WS only differentiated faces with changes in figures from those with changes in configuration and model faces. The latter two types of face yielded no difference. In the right hemisphere, TD controls showed differences in figure-changed faces from configure-changed and model faces. No difference was observed between the latter two face types. People with WS processed face changes in figures and configures distinct from model faces. These findings suggest that people with WS were not sensitive to the differentiation of configure-changed faces from model faces in the left hemisphere; moreover, people with WS processed configure-changed faces using a figure-changed strategy. Atypical neurological brainwaves emerged, given the normal behavioral patterns of people with WS. Another study by Grice et al. [31] on basic-level visual processing revealed atypical perceptual patterns of people with WS compared to TD controls. Participants viewed three types of images and recorded the corresponding brainwaves using event-related potentials: packmen, real squares, and Kanizsa squares. The results showed that people with WS processed illusory Kanizsa squares and packmen induced more negative brainwaves compared to real squares, whereas

TD controls processed the three types of images significantly differently. Kanizsa squares were processed more negatively than real squares, which, in turn, were processed more negatively than packmen by the TD. These findings suggest that people with WS are atypical in basic-level visual perception, leading to impaired integration of high-level cognitive information. This study is also compatible with the study by Mills et al. [32], both confirming atypical processing of people with WS to faces. In summary, people with WS exhibit asymmetrical performances between the brain and behavior in the nonverbal domain, similar to the patterns shown in the verbal domain.

5. Rehabilitation potential in people with Williams syndrome

Based on the findings of asymmetry of the brain and behavior in verbal and nonverbal studies on people with WS, our laboratory has started to explore their rehabilitation potential using advanced technology such as animation clips and virtual reality. The rationale for taking the theory of mind study as the pioneering work to examine the potential was to figure out the interaction of verbal and nonverbal abilities in people with WS. Theory of mind is an important cognitive ability that uses linguistic cues to execute reasoning in others' minds [33]. Several pioneering studies have been conducted on people with WS and DS in our laboratory. In the study by Hsu and Rao [23] on theory of mind to examine whether their ability in mental inference could be enhanced *via* presentation of animations, two studies were conducted on first-order and second-order mentalizing abilities. Traditionally, theory of mind tests were conducted using 2D black and white paper; in our study, animations with 3D design were tested. It was hypothesized that animations could attract the attention of people with WS more and induce their mentalizing abilities more deeply and easily. The classic study on theory of mind using the story of Sally and Anne [24] presented in animations was initially tested on TD groups of 3- and 4-year-olds for the first-order mentalizing test and on TD groups of 5- and 6-year-olds for the second-order mentalizing test. People with WS were tested later. The results confirmed that, compared to the critical age of 4 years in passing the first-order theory of mind, in Hsu and Rao [23], people with WS passed the test at 5.9 years old by using computerized animations. This finding confirms that computerized animations are a useful tool for successfully inducing the mentalizing ability of people with WS compared to previous studies of worse performance at 9–23 years old or unsuccessful passing age at 17.11 years old [10, 34, 35].

In the second-order theory of mind study using animations, people with WS passed the test at 7.14 years old mentally or 10.5 years old chronologically [26], which was closer to the critical age of 6–7 years old when passing the test. This finding in age was earlier than previous studies on the second-order study by Karmiloff-Smith et al. [10] at 9 years of age and by Sullivan and Tager-Flusberg [36] at 8.56 years of age. Both studies on theory of mind in people with WS revealed that people with WS were enhanced in mentalizing others' minds from another protagonist's viewpoint using animation clips. Further studies investigating knowledge of lexical semantics *via* advanced technology using virtual reality revealed worse performance in people with WS in a spatial navigation task [24]. In the spatial navigation task, people with WS showed similar practice effects in the supermarket shopping task from the first to the fourth navigation as their TD controls. This effect also sheds light on the rehabilitation potential of people with WS.

6. Conclusion

The asymmetry of local focusing and global ignorance of people with WS leads to impaired integration of contextual information in verbal and nonverbal domains. Importantly, some studies have shown normal behavioral performances given deficient underlying neurological mechanisms in people with WS, which is termed asymmetry of the brain and behavioral phenomena. Based on the findings of asymmetry, rehabilitation is possible for people with WS, who may use atypical neurological reflections to achieve TD behaviors. The brains of people with WS show plasticity in processing cognitive abilities across verbal and nonverbal domains, as neuroconstructivism claims [2, 3]. The brains of people with WS are shaped developmentally and dynamically. Advanced technologies in modern society, such as virtual reality, augmented reality, animation, and robotics, can be developed as rehabilitated tools for people with WS in the near future.

Acknowledgements

The corresponding author would like to thank all the supporting grants to the corresponding author, including the Fundamental Research Funds of the Central Universities in China for the startup funding received from Hunan University (531118010366), National Social Science Fund of China (22FYYB033), and partial support from Hunan University (521119400434, 521119400596).

Author details

Ching-Fen Hsu^{1,2*}, Qian Jiang^{2,3}, Shi-Yu Rao^{2,4}, Xiu-Zhi Fu^{2,3} and Xin Zheng^{2,3}

1 School of Foreign Languages and Literature, Wuhan University, China

2 Laboratory for Language Pathology and Developmental Neurosciences, Wuhan University, China

3 School of Foreign Languages, Hunan University, China

4 School of Foreign Languages, Changsha University, China

*Address all correspondence to: 00036126@whu.edu.cn

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Reed VA. An Introduction to Children with Language Disorders. 5th ed. New York: Pearson Communication Sciences and Disorders; 2018
- [2] Mareschal D, Johnson M, Sirois S, Spratling M, Thomas M, Westermann G. Neuroconstructivism I: How the Brain Constructs Cognition (Developmental Cognitive Neuroscience). New York: Oxford University Press; 2007
- [3] Sirois S, Westermann G, Mareschal D. Neuroconstructivism II: Perspectives and Prospects (Developmental Cognitive Neuroscience). New York: Oxford University Press; 2007
- [4] Westermann G, Mareschal D, Johnson MH, Sirois S, Spratling MW, Thomas MSC. Neuroconstructivism. *Developmental Science*. 2007;**10**(1):75-83
- [5] Karmiloff-Smith A. Williams syndrome. *Current Biology*. 1996;**17**(24):R1035-R1036
- [6] Morris CA. Introduction: Williams syndrome. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. 2010;**154C**(2):203-208
- [7] Korenberg JR, Chen XN, Hirota H, Lai L, Bellugi U, Burian D, et al. Genome structure and cognitive map of Williams syndrome. *Journal of Cognitive Neuroscience*. 2000;**1**:89-107
- [8] Donnai D, Karmiloff-Smith A. Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics*. 2000;**97**(2):164-171
- [9] Bellugi U, Lichtenberger L, Jones W, Lai Z, George M. The neurocognitive profile of Williams syndrome: A complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*. 2000;**12**:7-29
- [10] Karmiloff-Smith A, Klima E, Bellugi U, Grant J, Baron-Cohen S. Is there a social module? Language, face processing, and theory of mind in individuals with Williams syndrome. *Trends in Cognitive Sciences*. 1995;**7**:196-208
- [11] Hsu CF. Cross-modal contextual coherence of semantic integration in people with Williams syndrome. *Research in Developmental Disabilities*. 2013a;**34**(12):4319-4327
- [12] Hsu CF. Is the contextual effect weak in people with Williams syndrome? An investigation of information integration ability using pictures. *Research in Developmental Disabilities*. 2013b;**34**(3):932-939
- [13] Hsu CF. Modality effect of contextual integration in people with Williams syndrome. *Research in Developmental Disabilities*. 2014;**35**(7):1571-1578
- [14] Hsu CF, Lv P. Investigating basic emotion knowledge of people with Williams syndrome. *Research in Developmental Disabilities*. 2022;**129**:104308
- [15] Hsu CF. Contextual effects on figurative processing of nouns and verbs in people with Williams syndrome. *Research in Developmental Disabilities*. 2023;**134**:104429
- [16] Hsu CF, Tzeng JL. Contextual effect in people with Williams syndrome. *Research in Developmental Disabilities*. 2011;**32**(2):781-787

- [17] Hsu CF. Contextual integration of causal coherence in people with Williams syndrome. *Research in Developmental Disabilities*. 2013c;**34**(10):3332-3342
- [18] Hsu CF. Neural correlates of causal inferences and semantic priming in people with Williams syndrome: An fMRI study. *Journal of Intellectual Disabilities-Diagnosis and Treatment*. 2020a;**8**(4):698-709
- [19] Hsu CF. Understanding emotion language through events and words in people with Williams syndrome. *Child Neuropsychology*. 2020b;**26**(8):1091-1111
- [20] Hsu CF, Lv P. Emotion priming in people with Williams syndrome. *Brain Sciences*. 2023;**13**:467
- [21] Hsu CF, Chen JY. Deviant neural correlates of configural detection in facial processing of people with Williams syndrome. *Bulletin of Special Education*. 2014;**39**(1):61-84
- [22] Hsu CF, Karmiloff-Smith A, Tzeng O, Chin RT, Wang HC. Semantic knowledge in Williams syndrome: Insights from comparing behavioural and brain processes in false memory tasks. In: *Proceedings of the 6th International Conference on Development and Learning*. Vol. 6. London, UK: IEEE; 2007. pp. 48-52
- [23] Hsu CF, Rao SY. Computerized false belief tasks impact mentalizing ability in people with Williams syndrome. *Brain Sciences*. 2023;**13**:722
- [24] Hsu CF, Jiang Q, Rao SY. (preprint). Impact of virtual reality on memory in people with Williams syndrome. 10.20944/preprints202410.1809.v1
- [25] Hsu CF, Rao SY, Jiang Q. (preprint). Computerized second-order mentalization in people with Williams syndrome. 10.20944/preprints202410.1953.v1
- [26] Karmiloff-Smith A. Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences*. 1998;**2**(10):389-398
- [27] Thomas MSC, Baughman FD. Neuroconstructivism: Understanding typical and atypical developmental trajectories. *Enfance*. 2014;**3**:205-236
- [28] Rossen M, Klima ES, Bellugi U, Bihle A, Jones W. Interaction between language and cognition: Evidence from Williams syndrome. In: Beitchman JH, Cohen N, Konstantareas M, Tannock R, editors. *Language Learning and Behaviour*. New York, NY: Cambridge University Press; 1996. pp. 367-392
- [29] Tyler LK, Karmiloff-Smith A, Voice JK, Stevens T, Grant J, Udwin O, et al. Do individuals with Williams syndrome have bizarre semantics? Evidence for lexical organization using an on-line task. *Cortex*. 1997;**33**(3):515-527
- [30] Gagliardi C, Frigerio E, Burt DM, Gazzaniga I, Perrett D, Borgatti R. Facial expression recognition in Williams syndrome. *Neuropsychologia*. 2003;**41**(6):733-738
- [31] Grice SJ, de Haan M, Halit H, Johnson MH, Csibra G, Grant J, et al. ERP abnormalities of illusory contour perception in Williams syndrome. *Neuroreport*. 2003;**14**(14):1773-1777
- [32] Mills DL, Alvarez TD, George MS, Appelbaum LG, Bellugi U, Neville H. Electrophysiological studies of face processing in Williams syndrome. *Journal of Cognitive Neuroscience*. 2000;**12**(1):47-64
- [33] Frith U. *Autism: Explaining the Enigma*. Oxford: Blackwell Publishing; 1989

[34] Tager-Flusberg H, Boshart J, Baron-Cohen S. Reading the windows to the soul: Evidence of domain-specific sparing in Williams syndrome. *Journal of Cognitive Neuroscience*. 1998;**10**:631-639

[35] Tager-Flusberg H, Sullivan K. A componential view of theory of mind: Evidence from Williams syndrome. *Cognition*. 2000;**76**:59-89

[36] Sullivan K, Tager-Flusberg H. Second-order belief attribution in Williams syndrome: Intact or impaired? *American Journal of Mental Retardation*. 1999;**104**(6):523-532

iPS Cell Models of Human Single Gene Inheritance Diseases

Yuqing Ran and Xiaoling Guo

Abstract

Single gene inheritance diseases account for a large proportion of human congenital genetic diseases. For a long time, it has been difficult for scientists to directly obtain patient's somatic cells to establish *in vitro* cell models to study disease pathogenesis or screen personalized treatment drugs. Theoretically, induced pluripotent stem (iPS) cells have the potential to differentiate into various types of cells derived three germ layers such as cardiomyocytes, Leydig cells, RPE, podocytes, islet cells, neurons, et al. The advent of reprogramming a variety of somatic cells into iPS cells provided the possibility to generate different human cell models carrying the patient's genetic background and mutational profiles even with single gene inheritance disease phenotype to used for studying the pathological mechanisms, drug screening, or gene therapy *in vitro*. Here, we will discuss the development of iPS cells and its applications in single gene inheritance diseases, including key advancements, challenges, and future research directions.

Keywords: induced pluripotent stem (iPS) cells, human single gene inheritance diseases, reprogramming, cell models, application

1. Introduction

Single gene inherited diseases are a class of disorders caused by mutations in one gene, and have well-defined genetic mechanisms and phenotypic characteristics [1, 2]. Elucidating the pathological mechanisms of human single gene inherited diseases plays a key role in the discovery of new therapeutic strategies. The traditional animal models of transgenic mutations are invaluable for simulating human diseases to explore the potential pathological mechanisms at different developmental stages and specific cell types *in vivo*. However, due to species differences, animal models cannot truly reflect the clinical characteristics of disease [3, 4]. Moreover, it is generally difficult to obtain human end-terminal differentiated cells such as cardiomyocytes, retinal pigment epithelium (RPE), neuron, Leydig cells, and so on for *in vitro* pathological mechanism study [5, 6].

The emergence of induced pluripotent stem (iPS) cells breaks this limitation and becomes an attractive alternative because of their ability to be derived from somatic cells of patients with single gene inheritance diseases through cell reprogramming technology [7, 8]. At the same time, iPS cells with the patient's genetic background have self-renewal and differentiation potential, so they can theoretically be infinitely

induced to differentiate into any type of human cells to establish personalized disease cell models carrying single gene inheritance disease phenotype for precision medicine [9, 10].

Currently, both human embryonic stem (ES) cells and iPS cells have been used to model human genetic diseases. Although the initial *in vitro* cell models were developed using ESCs [11], there has been a gradual shift in preference toward iPS cells due to their lack of similar ethical issues associated with ES cells as well as the availability of iPS cells.

iPS cells are very similar to embryonic stem (ES) cells in that they both express pluripotency markers, and both have the potential to differentiate into three germ layers (endoderm, mesoderm, and ectoderm) [12–14]. Moreover, iPS cells can also be used to address the problems of immune rejection faced by ES cells [15, 16]. So, iPS cell models of human single gene inheritance diseases provide a powerful platform for studying disease mechanisms, drug screening, and gene therapy development.

2. Development of iPS cell technology

Since 2006, iPS reprogramming technology has developed rapidly, ushering in an exciting new era in the field of regenerative medicine. In 2006, Takahashi and Yamanaka achieved the first reprogramming iPS cells derived from mouse skin fibroblasts by introducing four key transcription factors Oct4, Sox2, Klf4, and c-Myc [17]. Subsequently, human somatic cells were also successfully reprogrammed into iPS cells using the same four transcription factors [14]. In 2008, Dunsbergen et al. [18] and Wernig et al. [19] demonstrated that iPS cells can also be produced using only three transcription factors (*Oct4*, *c-Myc*, *Klf4* or Oct4, Sox2, Klf4). Kim et al. [20] and Huangfu et al. [21] further improved reprogramming technology, showing that neural stem cells need only two transcription factors (Oct4 binds to Klf4, c-Myc, or Sox2) to be reprogrammed into iPS cells. Surprisingly, Kim et al. [22] successfully generated iPS cells derived from neural stem cells using only one transcription factor Oct4.

Initially, retroviruses vectors were widely used for delivering Yamanaka factors [7, 13, 14, 16]. This delivering method was highly efficient but had significant disadvantages, including the possibility of insertional mutagenesis to cause cancer risk. Subsequently, advances in reprogramming technology sought safer and more efficient methods. One such advancement was the use of Sendai virus (SeV), which was a non-integrating RNA virus [23]. Once, we successfully reprogrammed patient-derived urine cells into human iPS cells using Yamanaka factors such as SOX2, OCT4, KLF4, and c-MYC mediated by Sendai virus [24]. With the development of study, researchers had paid attention to the reprogramming with non-genetic approach, such as using RNA, proteins, and small molecules to reprogram patient somatic cells into iPS cells [25].

Yakubov et al. firstly [26] reprogrammed somatic cells into iPSCs using RNA-based methods. Warren et al. also [27] achieved reprogramming of iPS cells from human somatic cells by transducing mRNA encoding four transcription factors and enhancing mRNA stability through modifications like glycosylation and methylation at the 5' end. Anokye-Danso et al. [28] cloned exogenous miR302/367 genes into the pLOVE vector, and successfully reprogrammed human and mouse somatic cells into iPS cells by transfection vector. In addition, the recombinant proteins such as KLF4, OCT4, SOX2, and c-MYC (RPTs) were used to generate iPS cells [29–31].

The reprogramming of iPS cells involved acetylation, demethylation mediated chromatin remodeling, and some key signaling pathways such as MAPK, Wnt, TGF- β , MEK-ERK, and so on [32–34]. Therefore, iPS reprogramming can be achieved by intervening in epigenetics and key signaling pathways. Li et al. [35] demonstrated that iPS cells could be reprogrammed from somatic cells using *Oct4* and a combination of small molecules (VPA, CHIR99021, 616,452, and tranylcypromine), replacing *Sox2*, *Klf4*, and *c-Myc*. Hou et al. [36] further expanded this approach by replacing 4TFs with six small molecules (VPA, CHIR99021, 616,452, tranylcypromine, Forskolin, and DZNep) to achieving iPS cells. These methods can safely reprogram somatic cells into iPS cells without altering the host genome, opening new avenues for iPS cells application in the field of regenerative medicine.

These iPS cells, which are characterized by pluripotency and self-renewal capacity, must be induced by carefully optimized differentiation protocols to produce functional cell types, which were related to disease modeling and therapeutic applications. The differentiation processes involved mimicking natural developmental signals during embryonic development, using a combination of growth factors, signaling molecules, and environmental conditions to drive iPS cells toward the desired developmental lineage [37]. Using different small molecules, researchers could induce iPS cells differentiation into different specific cell types, such as neurons, cardiomyocytes, or liver cells, which were crucial for studying potential pathological mechanisms of diseases and developing targeted therapy drugs. Our team successfully used molecular compounds to induce human iPS cells differentiation into Leydig-like cells [38].

What's more, the emergence of CRISPR-Cas9 genome editing technology has further promoted the development of iPS cells. CRISPR-Cas9 is able to precisely and efficiently manipulate the genome and allows the introduction or correction of interest mutations in iPS cells. In addition, the gene processing of iPS cells holds important prospects for the development of personalized gene therapy.

3. Applications of iPS cell models in single gene inheritance diseases

Single gene inherited diseases are caused by mutations in a single gene [1, 2]. The pattern of these diseases typically follows the Mendelian inheritance patterns, including dominant, recessive, or X-linked inheritance [39, 40]. Although the pathogenesis and clinical manifestations of each single gene inherited disease are different, the etiology is clear and usually caused by mutation in one gene. The construction of disease models from iPS cells provides a powerful tool for in-depth understanding of the pathological mechanisms of these diseases and the development of new treatment strategies. The following are specific application examples of iPS cell models in several single gene inherited diseases.

3.1 iPS cell models in Alzheimer's disease (AD)

AD, a fatal condition, is a neurological disease characterized by progressively declining cognitive processes, such as memory and learning, and irreversible neuro degeneration [41]. Currently, the main treatment for AD is symptomatic therapy, such as using acetylcholinesterase inhibitors [42] and NMDA receptor antagonists [43] to alleviate symptoms, but the disease cannot be cured or reversed.

Kondo T et al. [44] generated iPS cells from somatic cells of patients with familial and sporadic AD and differentiated them into neural cells. Neurons derived from

patient iPS cells with a familial amyloid precursor protein (APP)-E693D mutation leads to endoplasmic reticulum (ER) and oxidative stress. Docosahexaenoic acid (DHA) could alleviate the stress responses in AD neural cells derived iPS cells, suggesting that DHA may in fact be effective for a subset of patients. This further demonstrates the utility of patient-specific iPS cells in investigating the underlying mechanisms of AD and evaluating the effects of drug therapy.

3.2 iPS cell models in Noonan syndrome

First reported in 1968, Noonan Syndrome (NS), also named LEOPARD Syndrome, is a relatively common autosomal dominant disorder [45]. The incidence of NS ranges from approximately 1:1000 to 2500 and is characterized by distinct craniofacial features, short stature, congenital cardiac anomalies, developmental delays, and variable neurocognitive impairment [46–49]. Approximately 80% of patients with NS harbor gene mutations, which are involved in the RAS/mitogenactivating protein kinase (MAPK) pathway [48, 50]. The gene of serine/threonine kinase rapidly accelerated fibrosarcoma 1 (RAF1) is crucial in the conserved RAS/MAPK pathway. NS can be caused by the gene mutations in RAF1 [48], which may account for about 3–17% patients with NS [51].

Youngee et al. [52] used skin fibroblasts cells from patients with Noonan syndrome to reprogram into iPS cells. Subsequently, these iPS cells were induced to differentiate into neural progenitor cells and neurospheres, and further differentiated into neurons and glial cells. The results showed that mutations in SHP2 lead to premature and excessive generation of glial cells during neurodevelopment, revealing the cellular and molecular basis of neurodevelopmental abnormalities in Noonan syndrome. This approach not only preserves the patient's genetic mutation background but also accurately simulates the cellular and molecular mechanisms of the disease. It also provides a powerful model system for understanding the pathology of Noonan syndrome, which could help to develop potential treatment strategies.

3.3 iPS cell models in Lowe syndrome

First described by Lowe et al. in 1952 [53], Lowe syndrome (LS), an inherited rare X-linked multisystemic disorder syndrome with the prevalence of approximately 1 per 500,000, is characterized by congenital cataracts, neonatal hypotonia, intellectual impairment, and renal tubular dysfunctions with slowly progressive renal failure [54]. LS can mainly affect the function of the eye, brain, and kidney and is caused by the gene mutations of OCRL at Xq25-26, which encodes an inositol polyphosphate 5-phosphatase (IPP-5P) in the trans-Golgi network [54, 55]. IPP-5P is essential for proper cell migration, spreading, cytoskeleton-plasma membrane adhesion, and various protein-protein interactions [56].

Akhtar et al. [57] extracted skin fibroblasts cells from patients with Lowe syndrome and reprogrammed these cells into iPS cells. Subsequently, they induced these iPS cells to differentiate into neuronal cells. Based on these neuronal cell models, they observed abnormal phospholipid metabolism caused by mutations in the *OCRL1* gene, which in turn affects the neural development process. These results indicate that the use of patient-derived iPS cell models can effectively simulate the neurodevelopmental deficits of Lowe syndrome, and provide a deeper understanding of the pathological mechanisms of the disease.

4. Conclusions

iPS cell model has demonstrated significant potential and value in the study of human single gene inheritance diseases. By reprogramming the patient's somatic cells into iPS cells and further differentiating them into disease-specific cell types, researchers can simulate key processes and phenotypic characteristics of disease development. However, iPS cell model still faces many challenges, such as the efficiency and consistency of cell differentiation, the complexity of the model, and the feasibility of translating research findings into clinical applications.

Future research directions may include optimizing the differentiation methods of iPS cell, improving their accuracy in simulating disease processes, and developing more targeted treatment strategies. In addition, combining gene editing technology and high-throughput screening methods to establish a larger scale iPS cell resource library will help accelerate the development of personalized treatment methods.

In summary, iPS cells provide a powerful experimental platform for studying human single gene inheritance diseases, and also open broad prospects for future scientific research and clinical treatment.

Acknowledgements

This work was supported by the Natural Science Foundation of Zhejiang Province (LY24H020008), and the Fourth Batch of Wenzhou Medical University "Outstanding and Excellent Youth Training Project" (604090352/640).

Conflict of interest

The authors have declared that no competing interests exist.

Abbreviations

iPS	induced pluripotent stem
RPE	retinal pigment epithelium
ES	embryonic stem
ESCs	embryonic stem cells
SeV	sendai virus
AD	Alzheimer's disease
ER	endoplasmic reticulum
DHA	docosahexaenoic acid
NS	Noonan syndrome
MAPK	mitogenactivating protein kinase
RAF1	rapidly accelerated fibrosarcoma
LS	Lowe syndrome
IPP-5P	inositol polyphosphate 5-phosphatase

Author details

Yuqing Ran¹ and Xiaoling Guo^{1,2,3*}

1 Basic Medical Research Center, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

2 Department of Scientific Research, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

3 Key Laboratory of Structural Malformations in Children, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China

*Address all correspondence to: guoxling@hotmail.com

IntechOpen

© 2024 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Weatherall DJ. Single gene disorders or complex traits: Lessons from the thalassaemias and other monogenic diseases. *BMJ*. 2000;**321**(7269):1117-1120. DOI: 10.1136/bmj.321.7269.1117
- [2] Aksentijevich I, Schnappauf O. Molecular mechanisms of phenotypic variability in monogenic auto-inflammatory diseases. *Nature Reviews Rheumatology*. 2021;**17**(7):405-425. DOI: 10.1038/s41584-021-00614-1
- [3] Jucker M. The benefits and limitations of animal models for translational research in neurodegenerative diseases. *Nature Medicine*. 2010;**16**(11):1210-1214. DOI: 10.1038/nm.2224
- [4] Doncheva NT, Palasca O, Yarani R, Litman T, Anthon C, Groenen MAM, et al. Human pathways in animal models: Possibilities and limitations. *Nucleic Acids Research*. 2021;**49**(4):1859-1871. DOI: 10.1093/nar/gkab012
- [5] Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology*. 2009;**27**(3):275-280. DOI: 10.1038/nbt.1529
- [6] Itskovitz-Eldor J, Schuldiner M, Karsenti D, Eden A, Yanuka O, Amit M, et al. Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Molecular Medicine*. 2000;**6**(2):88-95
- [7] Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, et al. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature*. 2008;**451**(7175):141-146. DOI: 10.1038/nature06534
- [8] Ye Z, Zhan H, Mali P, Dowey S, Williams DM, Jang YY, et al. Human-induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. *Blood*. 2009;**114**(27):5473-5480. DOI: 10.1182/blood-2009-04-217406
- [9] Marchetto MC, Winner B, Gage FH. Pluripotent stem cells in neurodegenerative and neurodevelopmental diseases. *Human Molecular Genetics*. 2010;**19**(R1):R71-R76. DOI: 10.1093/hmg/ddq159
- [10] Zhu H, Lensch MW, Cahan P, Daley GQ. Investigating monogenic and complex diseases with pluripotent stem cells. *Nature Reviews. Genetics*. 2011;**12**(4):266-275. DOI: 10.1038/nrg2951
- [11] Ben-David U, Kopper O, Benvenisty N. Expanding the boundaries of embryonic stem cells. *Cell Stem Cell*. 2012;**10**(6):666-677. DOI: 10.1016/j.stem.2012.05.003
- [12] Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. Embryonic stem cell lines derived from human blastocysts. *Science*. 1998;**282**(5391):1145-1147. DOI: 10.1126/science.282.5391.1145
- [13] Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007;**318**(5858):1917-1920. DOI: 10.1126/science.1151526
- [14] Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from

adult human fibroblasts by defined factors. *Cell*. 2007;**131**(5):861-872. DOI: 10.1016/j.cell.2007.11.019

[15] Moradi S, Mahdizadeh H, Šarić T, Kim J, Harati J, Shahsavarani H, et al. Research and therapy with induced pluripotent stem cells (iPSCs): Social, legal, and ethical considerations. *Stem Cell Research and Therapy*. 2019;**10**(1):341. DOI: 10.1186/s13287-019-1455-y

[16] Wang Z, Zheng J, Pan R, Chen Y. Current status and future prospects of patient-derived induced pluripotent stem cells. *Human Cell*. 2021;**34**(6):1601-1616. DOI: 10.1007/s13577-021-00592-2

[17] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;**126**(4):663-676. DOI: 10.1016/j.cell.2006.07.024

[18] Duinsbergen D, Eriksson M, T Hoen PA, Frisén J, Mikkers H. Induced pluripotency with endogenous and inducible genes. *Experimental Cell Research*. 2008;**314**(17):3255-3263. DOI: 10.1016/j.jyexcr.2008.06.024

[19] Wernig M, Zhao JP, Pruszak J, Hedlund E, Fu D, Soldner F, et al. Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;**105**(15):5856-5861. DOI: 10.1073/pnas.0801677105

[20] Kim JB, Zaehres H, Wu G, Gentile L, Ko K, Sebastiano V, et al. Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature*. 2008;**454**(7204):646-650. DOI: 10.1038/nature07061

[21] Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, et al. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nature Biotechnology*. 2008;**26**(11):1269-1275. DOI: 10.1038/nbt.1502

[22] Kim JB, Sebastiano V, Wu G, Araújo-Bravo MJ, Sasse P, Gentile L, et al. Oct4-induced pluripotency in adult neural stem cells. *Cell*. 2009;**136**(3):411-419. DOI: 10.1016/j.cell.2009.01.023

[23] Miyamoto K, Akiyama M, Tamura F, Isomi M, Yamakawa H, Sadahiro T, et al. Direct In vivo reprogramming with Sendai virus vectors improves cardiac function after myocardial infarction. *Cell Stem Cell*. 2018;**22**(1):91-103.e5. DOI: 10.1016/j.stem.2017.11.010

[24] Qian R, Wang L, Chen H, Ding Y, Wang D, Chen C, et al. Establishment of a human induced pluripotent stem cell line (WMUi031-a) from a Lowe syndrome patient carrying a OCRL gene mutation (c.2626dupA). *Stem Cell Research*. 2021;**53**:102294. DOI: 10.1016/j.scr.2021.102294

[25] Guo XL, Chen JS. Research on induced pluripotent stem cells and the application in ocular tissues. *International Journal of Ophthalmology*. 2015;**8**(4):818-825. DOI: 10.3980/j.issn.2222-3959.2015.04.31

[26] Yakubov E, Rechavi G, Rozenblatt S, Givol D. Reprogramming of human fibroblasts to pluripotent stem cells using mRNA of four transcription factors. *Biochemical and Biophysical Research Communications*. 2010;**394**(1):189-193. DOI: 10.1016/j.bbrc.2010.02.150

[27] Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, et al. Highly efficient reprogramming to pluripotency and directed differentiation of human

cells with synthetic modified mRNA. *Cell Stem Cell*. 2010;7(5):618-630. DOI: 10.1016/j.stem.2010.08.012

[28] Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, et al. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*. 2011;8(4):376-388. DOI: 10.1016/j.stem.2011.03.001

[29] Zhou H, Wu S, Joo JY, Zhu S, Han DW, Lin T, et al. Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell*. 2009;4(5):381-384. DOI: 10.1016/j.stem.2009.04.005

[30] Nemes C, Varga E, Polgar Z, Klincumhom N, Pirity MK, Dinnyes A. Generation of mouse induced pluripotent stem cells by protein transduction. *Tissue Engineering. Part C, Methods*. 2014;20(5):383-392. DOI: 10.1089/ten.TEC.2013.0026

[31] Kim D, Kim CH, Moon JI, Chung YG, Chang MY, Han BS, et al. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell*. 2009;4(6):472-476. DOI: 10.1016/j.stem.2009.05.005

[32] Peng J, Zhang WJ, Zhang Q, Su YH, Tang LP. The dynamics of chromatin states mediated by epigenetic modifications during somatic cell reprogramming. *Frontiers in Cell and Development Biology*. 2023;11:1097780. DOI: 10.3389/fcell.2023.1097780

[33] Jiang D, Li T, Guo C, Tang TS, Liu H. Small molecule modulators of chromatin remodeling: From neurodevelopment to neurodegeneration. *Cell and Bioscience*. 2023;13(1):10. DOI: 10.1186/s13578-023-00953-4

[34] Huang J, Chen T, Liu X, Jiang J, Li J, Li D, et al. More synergetic cooperation of

Yamanaka factors in induced pluripotent stem cells than in embryonic stem cells. *Cell Research*. 2009;19(10):1127-1138. DOI: 10.1038/cr.2009.106

[35] Li Y, Zhang Q, Yin X, Yang W, Du Y, Hou P, et al. Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell Research*. 2011;21(1):196-204. DOI: 10.1038/cr.2010.142

[36] Hou P, Li Y, Zhang X, Liu C, Guan J, Li H, et al. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science*. 2013;341(6146):651-654. DOI: 10.1126/science.1239278

[37] Gao X, Li R, Cahan P, Zhao Y, Yourick JJ, Sprando RL. Hepatocyte-like cells derived from human induced pluripotent stem cells using small molecules: Implications of a transcriptomic study. *Stem Cell Research and Therapy*. 2020;11(1):393. DOI: 10.1186/s13287-020-01914-1

[38] Chen X, Li C, Chen Y, Xi H, Zhao S, Ma L, et al. Differentiation of human induced pluripotent stem cells into Leydig-like cells with molecular compounds. *Cell Death and Disease*. 2019;10(3):220. DOI: 10.1038/s41419-019-1461-0

[39] Berkovic SF, Scheffer IE. Epilepsies with single gene inheritance. *Brain and Development*. 1997;19(1):13-18. DOI: 10.1016/s0387-7604(96)00060-5

[40] Jedidi I, Ouchari M, Yin Q. Sex chromosomes-linked single-gene disorders involved in human infertility. *European Journal of Medical Genetics*. 2019;62(9):103560. DOI: 10.1016/j.ejmg.2018.10.012

[41] Tarawneh R, Holtzman DM. The clinical problem of symptomatic Alzheimer disease and mild

cognitive impairment. Cold Spring Harbor Perspectives in Medicine. 2012;**2**(5):a006148. DOI: 10.1101/cshperspect.a006148

[42] Birks J, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. Cochrane Database of Systematic Reviews. 2006;**25**(1):CD001190. DOI: 10.1002/14651858.CD001190.pub2. Update in: Cochrane Database Syst Rev. 2018; 6:CD001190. DOI: 10.1002/14651858.CD001190

[43] Reisberg B, Doody R, Stöffler A, Schmitt F, Ferris S, Möbius HJ, et al. Memantine in moderate-to-severe Alzheimer's disease. The New England Journal of Medicine. 2003;**348**(14):1333-1341. DOI: 10.1056/NEJMoa013128

[44] Kondo T, Asai M, Tsukita K, Kutoku Y, Ohsawa Y, Sunada Y, et al. Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular A β and differential drug responsiveness. Cell Stem Cell. 2013;**12**(4):487-496. DOI: 10.1016/j.stem.2013.01.009

[45] Noonan JA. Hypertelorism with turner phenotype. A new syndrome with associated congenital heart disease. American Journal of Diseases of Children. 1968;**116**(4):373-380. DOI: 10.1001/archpedi.1968.02100020377005

[46] Mendez HM, Opitz JM. Noonan syndrome: A review. American Journal of Medical Genetics. 1985;**21**(3):493-506. DOI: 10.1002/ajmg.1320210312

[47] Noonan JA, Noonan syndrome. An update and review for the primary pediatrician. Clinical Pediatrics (Phila). 1994;**33**(9):548-555. DOI: 10.1177/000992289403300907

[48] Roberts AE, Allanson JE, Tartaglia M, Gelb BD. Noonan syndrome.

Lancet. 2013;**381**(9863):333-342. DOI: 10.1016/S0140-6736(12)61023-X

[49] Tartaglia M, Gelb BD, Zenker M. Noonan syndrome and clinically related disorders. Best Practice and Research. Clinical Endocrinology and Metabolism. 2011;**25**(1):161-179. DOI: 10.1016/j.beem.2010.09.002

[50] Niihori T, Nagai K, Fujita A, Ohashi H, Okamoto N, Okada S, et al. Germline-activating RAS2 mutations cause Noonan syndrome. American Journal of Human Genetics. 2019;**104**(6):1233-1240. DOI: 10.1016/j.ajhg.2019.04.014

[51] Yin JC, Platt MJ, Tian X, Wu X, Backx PH, Simpson JA, et al. Cellular interplay via cytokine hierarchy causes pathological cardiac hypertrophy in RAF1-mutant Noonan syndrome. Nature Communications. 2017;**8**:15518. DOI: 10.1038/ncomms15518

[52] Ju Y, Park JS, Kim D, Kim B, Lee JH, Nam Y, et al. SHP2 mutations induce precocious gliogenesis of Noonan syndrome-derived iPSCs during neural development in vitro. Stem Cell Research and Therapy. 2020;**11**(1):209. DOI: 10.1186/s13287-020-01775-8

[53] Lowe CU, Terrey M, MacLachlan EA. Organic-aciduria, decreased renal ammonia production, hydrophthalmos, and mental retardation; a clinical entity. A.M.A. American Journal of Diseases of Children. 1952;**83**(2):164-184. DOI: 10.1001/archpedi.1952.02040060030004

[54] Loi M. Lowe syndrome. Orphanet Journal of Rare Diseases. 2006;**1**:16. DOI: 10.1186/1750-1172-1-16

[55] Allison SJ. Lowe syndrome: Dysregulation of autophagosome-lysosome fusion in Lowe syndrome.

Nature Reviews. Nephrology.
2016;**12**(9):507. DOI: 10.1038/
nrneph.2016.110

[56] Coon BG, Mukherjee D, Hanna CB, Riese DJ 2nd, Lowe M, Aguilar RC. Lowe syndrome patient fibroblasts display Ocr1-specific cell migration defects that cannot be rescued by the homologous Inpp5b phosphatase. Human Molecular Genetics. 2009;**18**(23):4478-4491. DOI: 10.1093/hmg/ddp407

[57] Akhtar BM, Bhatia P, Acharya S, Sharma S, Sharma Y, Bhuvanendran Nair Suseela Devi A, et al. A human stem cell resource to decipher the biochemical and cellular basis of neurodevelopmental defects in Lowe syndrome. Biology Open. 2022;**11**(1):bio059066. DOI: 10.1242/bio.059066

Comparative Analysis of Carcinogenic Compounds in Cooking Oil Emissions and Cigarette Smoke: Implications for Genetic Disorders and Cancer Susceptibility

*Niranjana Karukayil Gopalakrishnan, Sana Syed Ansar,
Manikantan Pappuswamy and Aditi Chaudhary*

Abstract

This chapter endeavours to elucidate the striking similarities between the emissions originating from cooking oils and the smoke produced by cigarettes, both teeming with a cocktail of carcinogenic compounds such as polycyclic aromatic hydrocarbons, aromatic amines, and nitro-polycyclic aromatic hydrocarbons. This analysis posits a profound potential for noteworthy chromosomal modifications akin to those observed in response to exposure to cigarette smoke. Moreover, an augmented prevalence of genetic disorders, encompassing spontaneous abortion, is envisaged. The principal thrust of this paper lies in elaborating upon the inherent cancer-inducing attributes of these carcinogens present in both cooking oil fumes and cigarette smoke. Furthermore, it delves into the demographic cohorts exhibiting heightened susceptibility to such cytogenetic alterations.

Keywords: cooking oil fumes, smoke point, polycyclic aromatic hydrocarbons, indoor air pollution, repeatedly heated cooking oils, lung cancer, particulate matter (PM), volatile organic compounds (VOCs)

1. Introduction

Cooking oil fumes (COFs) represent a complex array of substances, with their composition varying based on the cooking method (e.g., boiling, frying etc.) and the type of oil used (e.g., rapeseed oil, coconut oil, sunflower oil, soybean oil etc.). Additionally, the presence and efficiency of fume extractors play a significant role.

Research indicates that volatile particulates from lard heated to 200 and 300°C, as well as soybean oil heated to 300°C, exhibit mutagenic properties in the Salmonella/microsomal assay with S9 mix. According to epidemiological studies, women who are exposed to cooking oil fumes have a higher chance of developing lung cancer. Cooking practises and the utilisation of fume extractors are plausible biological determinants of lung cancer risk. They impact the concentration of known mutagens and carcinogens in oil fumes, including polycyclic aromatic hydrocarbons (PAHs), heterocyclic aromatic amines, benzene, and formaldehyde [1–3].

Deep-frying is a major contributor to indoor air pollution. Recently, commercial kitchens have begun using vegetable margarine for frying, which can lead to pulmonary inflammation that intensifies with prolonged exposure. Fast-food restaurants, popular for their convenience and affordability, predominantly serve fried foods. More people come into contact with cooking fumes and other pollution for longer periods of time as these establishments proliferate and consumer appetite for their products rises. High-temperature cooking generates carcinogenic fumes. Epidemiological studies indicate that cooks and bakers are at a heightened risk of developing cancer. Particularly, non-smoking women in East Asia exhibit a higher risk of cancer. Among bakers and pastry cooks, there is a notable excess of bronchial carcinoma, with a relative risk (RR) of 3.6 and a confidence interval (CI) of 1.3–10.4. This signifies that these individuals are 3.6 times more likely to develop bronchial carcinoma compared to a reference group, possibly the general population or another occupational group. The confidence interval underscores the statistical significance of this finding, indicating that the true relative risk is highly likely to fall within this range. Aldehydes present in cooking fumes can infiltrate cells, causing damage by reacting with macromolecules such as DNA [4–6].

Cooking oil fumes (COFs) from hot oil may play a role in the development of lung cancer. Since 2021, South Korea has recognised occupational lung cancer among individual cafeteria workers. Occupational exposure to COFs can influence lung cancer development, and ventilation is a critical factor. COFs can cause lung cancer in women, independent of smoking. Cooking emissions and inadequate kitchen ventilation are linked to increased lung cancer risks, especially in women. Furthermore, poor ventilation during cooking can elevate the risk of lung cancer [7–9].

2. The impact of various cooking oils and cooking methods on human health

Coronary heart disease (CHD) is the leading cause of death worldwide, with its occurrence rising swiftly, especially in developing nations like India. Edible oils play a critical role in the development, treatment, management, and prevention of CHD. Selecting the appropriate edible oil is particularly important in India, where cooking techniques differ from those in Western countries. Edible oils consist of different fatty acids, which can be classified into three categories: saturated fatty acids, monounsaturated fatty acids, and polyunsaturated fatty acids [10].

Saturated fatty acids are generally considered harmful because they can elevate total cholesterol (TC) and low-density lipoprotein (LDL) cholesterol, increasing the risk of atherosclerosis. According to research, women are more likely to prevent coronary heart disease by substituting unhydrogenated monounsaturated and polyunsaturated fats for saturated and trans fats than by consuming fewer lipids overall. Additionally, dietary fat intake has been directly linked to the risk of postmenopausal invasive breast cancer [11].

In Chinese cuisine, stir-frying, frying, and deep-frying are common cooking methods. We also measured the rate with which cooks employed each approach and asked about their typical cooking practises, such as waiting for cooking oil vapours to release before starting to cook. To preserve food as fresh as possible, traditional Chinese cooking emphasises rapid frying in oils heated to high degrees. Studies have shown that exposure to cooking oil fumes (COFs) among military cooks is higher than that among office workers. The levels of airborne particulate PAHs in kitchens significantly exceed those found in office environments [1].

Cooking generates various substances, including particulate matter, liquid oil droplets, water vapour, and gaseous pollutants. Cooking techniques include boiling food in soup or water, as in hot pots, are popular in China. This technique does not use oil at all and cooks' food at a lower temperature, but it does not always result in less pollution or harm. The type and shape of food show minimal impact on the particulate matter (PM) emission characteristics during boiling [12].

Boiling water releases more large particles, whereas heating oil releases more little particles. These are the differences in the emission characteristics of the two processes. Studies indicate that PM_{2.5} emissions generally decrease in the following order: deep-frying (0.709–2.731 mg/m³) > stir-frying (0.700–0.958 mg/m³) > stewing (~0.573 mg/m³) > quick-frying (0.140–0.433 mg/m³) > boiling (0.004–0.247 mg/m³) > steaming (0.011–0.088 mg/m³), with most levels exceeding the national indoor air standards.

The average concentration of PM_{2.5} in the immediate breathing zone during family cooking was found to be 0.599 mg/m³. Housewives, including some pregnant women who are particularly sensitive to PM_{2.5}, are typically the primary participants in home cooking. Unlike commercial restaurants, home cooking involves less intense heat and smaller quantities of food [13].

Although spending more time indoors can lessen the impacts of outdoor air pollution, it can also raise the chance of breathing in high concentrations of interior pollutants, like those from heating and cooking. Particulate matter emitted by heated cooking oil poses significant health hazards. Among indoor pollutants, particulate matter is a leading cause of various health issues, including heart attacks, severe asthma, reduced lung function, respiratory irritation, and breathing difficulties [14].

Inhaling cooking oil fumes is harmful to human health. Studies have shown that coconut and olive oil produce the most particles at low temperatures, while peanut and canola oil generate the most particles at high temperatures [15]. One of the main sources of indoor particulate matter, cooking oil fumes (COFs), are linked to respiratory symptoms such as coughing congestion in the nose, and diminished lung function. The exposure level to COFs during typical Chinese domestic cooking can cause short-term detrimental effects on cardiopulmonary health, trigger inflammatory responses, and induce oxidative stress [16].

Fats, alkanes, and olefins are the main constituents of oil fumes, which are then followed by aldehydes, ketones, esters, aromatic compounds, and heterocyclic compounds. Studies have shown that cooks in both commercial and domestic kitchens are exposed to extremely high levels of particulate and gaseous pollutants. Real-time measurements in a Chinese food store revealed that during cooking, the mass and number concentrations of fine particulate matter (PM_{2.5}) increased by factors of 12 and 85, respectively, compared to non-cooking periods [16]. Temperature is the most significant factor influencing the rate of evaporative emissions of cooking oil fumes, with the evaporation rate and pollutant emissions being highest from unrefined oil samples. The popularity of fried foods means that cooking procedures often involve

volatilising oils at high temperatures, making oil temperature a major factor in cooking emissions. Negative health outcomes from exposure to particulate emissions from cooking with vegetable oils include respiratory symptoms, reduced lung function, neonatal complications, and cardiovascular diseases [17].

Fast-food is becoming increasingly significant in society due to its quick availability and appealing taste. This trend increases the exposure of cooks to cooking oil fumes, which can have drastic long-term health effects [18]. In southern India, where coconut oil is commonly used for cooking, even low-temperature heating of oil produces particles that affect those working in home kitchens.

One of the most popular and sophisticated methods of food preparation is deep-frying, frequently practised in fast-food establishments. The quality of the oil or fats used to fry meals greatly influences their quality. For instance, deep-frying in sunflower oil raises the concentration of linoleic acid but only results in a 2% dry increase in total fat. γ - and α -linoleic acids are oxidised, whereas eicosapentaenoic and docosahexaenoic acids are retained. Additionally, deep-frying triggers the production of volatile chemicals, mainly aldehydes and ketones from polyunsaturated fatty acids, as well as free amino acids through proteolysis. Additionally, significant quantities of furanic compounds, such as furan and furfuryl alcohol, are generated during deep-frying, especially when frying coated fish [19].

A product that is fried tastes excellent, has a pleasant flavour, and can be made quickly. Emulsifiers, anti-polymerising agents, and natural and synthetic antioxidants improve the performance during frying [20]. During frying humidity and air are mixed with hot oil and cause chemical reactions in the oil used for frying. As a result, these fats and oils are affected by autoxidation, thermo-oxidation, pyrolysis, and thermal polymerisation as well as oxidative decomposition, which alter and degrade the composition of the oil thereby produce such as dimeric and polymeric materials which can be harmful to health [18, 20].

A number of factors, such as the type of frying oil used, the temperature, the length of time the food is fried, the frying pan, and the moisture level of the food, affect the kind, quantity, and creation of hazardous compounds produced during frying. An important consideration is the type of frying oil used, which serves as a medium for heat transfer [21, 22]. When oils are used for extended periods of time, they oxidise and produce hydroxides as well as volatile substances including aldehydes, ketones, carboxylic acids, and other unwanted substances.

The majority of human life is spent indoors, and cooking contributes significantly to indoor air pollution. Consequently, it is crucial to analyse how much human exposure occurs to home cooking fumes. For the quality of the frying medium and fried foods, as well as the shelf life of the food products, the choice and characteristics of the oils used in the frying process are crucial. Antioxidants can decrease the oxidation of frying oil, but their effectiveness diminishes at high frying temperatures. Lignan compounds in sesame oil are effective antioxidants in deep-fat frying [23]. Traditional oils high in polyunsaturated fatty acids (PUFA) are not ideal for frying due to their fatty acid composition, iodine value (IV) characteristics, and their higher oxidation tendency [24].

Vegetable oils are frequently heated to high temperatures in cooking. Cooking oils that have been heated repeatedly (RCO) can release a number of chemicals, some of which are known to be carcinogenic and include polycyclic aromatic hydrocarbons (PAHs).

The consumption of RCO and inhalation of its fumes are associated with a high incidence of genotoxic, mutagenic, and tumorigenic effects, as well as various

cancers. Cooking fumes have been found to increase the occurrence of aberrant cells, including chromosomal breaks, fragments, exchanges, multiple chromosomal damages, and micronuclei, in a dose-dependent manner [25].

One of the main health advantages of plant-based oils is the breakdown of important polyphenol antioxidants that occur during the heating process. Repeatedly heating fats, especially polyunsaturated fats, at high temperatures beyond their smoke point can result in the formation of carcinogenic compounds in both the oil and the foods cooked in it [25]. Unrefined oils typically have lower smoke points and can become rancid more quickly, so it is important to understand their smoke points and store them properly.

Studies suggest that the fumes from cooking oils containing PAHs and NPAHs may increase the risk of lung cancer among cooks [26]. The addition of natural antioxidants, such as catechin, could reduce the presence of these carcinogens [27]. Sixteen PAHs have been identified as being emitted through deep-frying and frying methods using rapeseed, soybean, peanut, and olive oils [28].

Cooking generates a variety of volatile, semi-volatile, and particulate pollutants. Since most people spend over 80% of their time indoors, indoor air quality has a significant impact on human health compared to outdoor environments [29]. One of the leading causes of lung cancer among women is exposure to kitchen cooking fumes. Indoor pollution associated with cooking oil increases lung cancer susceptibility in China, primarily due to the frequent use of deep-frying and poor kitchen ventilation in Chinese cooking traditions [30].

The presence of triacylglycerol (TAG) dimers in fresh cooking oils is typical because these dimers form during the deodorisation stage of oil refining, which involves high-temperature processing. These acyclic dimers can potentially lead to the creation of polyaromatic hydrocarbons or cyclic compounds [31].

The smoke point is a crucial physical parameter used to determine the stability of oils. It generally correlates with the accumulation of decomposition products. When oil starts to smoke, it marks the onset of both flavour and nutritional degradation [32]. A low smoke point usually indicates low thermal stability, while a higher smoke point suggests that the oil is of superior quality [33].

Higher smoke points are generally associated with monounsaturated oils, which include high oleic sunflower and safflower oils, canola oil, peanut oil, almond oil, olive oil, avocado oil, and so on. Sesame oil, on the other hand, contains nearly equal proportions of monounsaturated and polyunsaturated fats [34, 35]. **Table 1** defines list of smoke index of selected cooking oil types used.

3. The impact of ventilation in cooking areas on human health

Wood and charcoal are the main biomass fuels used by one-third of the world's population for cooking. There are an estimated 3.8 million premature deaths each year due to the household air pollution caused by burning these fuels, which is the primary environmental cause of mortality. Research indicates that ventilation improvements can reduce smoke concentration almost as effectively as many clean stove interventions implemented in real-world scenarios. Moreover, kitchen concentration is an inaccurate measure of exposure in households using biomass. This suggests that the interplay between stoves, fuels, housing conditions, and cooking behaviours must be considered when evaluating the success of clean cooking policies [40, 41].

The relationship between kitchen ventilation, cooking fuels, and cooking oil consumption and the risk of lung cancer was estimated using Cox proportional

S. no	Type of cooking oil	Smoke index	Reference
1.	Coconut oil	177°C	[36]
2.	Palm oil	235°C	[36]
3.	Sesame oil	177°C	[36]
4.	Olive oil	210°C	[36]
5.	Sunflower oil	227°C	[36]
6.	Rapeseed oil	107°C	[37]
7.	Peanut oil	232°C	[37]
8.	Avocado oil	271°C	[38]
9.	Butter	177°C	[38]
10.	Castor oil	200°C	[39]

Table 1.
Smoke index of selected cooking oil types used.

hazards regression and was reported as hazard ratios (HR) with a 95% confidence interval. Poor kitchen ventilation was associated with a 49% increased risk of lung cancer compared to adequate ventilation. Compared to women who were never exposed to inadequate ventilation, those who encountered it as children or adults had a 69% higher likelihood of lung cancer. Lung cancer accounts for approximately 18% of cancer deaths worldwide, with many of these deaths being preventable and largely attributable to tobacco use. However, Asian women have a relatively high rate of lung cancer despite a low prevalence of smoking [40, 42–45].

Chinese cooking techniques frequently release a variety of carcinogens into the air, which cooks—most of whom are women—inevitably inhale. While smoking tobacco is the primary risk factor for developing lung cancer, most Han Chinese women with lung cancer are non-smokers [46]. Lung cancer among non-smokers is a significant health issue that is gaining increasing recognition. Studies have shown that among lifetime non-smokers, lung cancer incidence is significantly higher in women than in men [47].

Genetic and environmental factors, particularly occupational exposure, can also elevate the risk of lung cancer. High levels of indoor particulate matter and other harmful indoor air pollutants are a factor in Taiyuan's increased lung cancer incidence rate among lifetime non-smokers. Research has found a significant interaction between cooking oil fume exposure and TERT (telomerase reverse transcriptase) polymorphism on lung cancer risk among Chinese non-smoking females [48].

Lung cancer is the leading cause of death among males in industrialised countries and is on the rise among females [49]. Cooks are exposed to a broad range of substances produced during the cooking and frying processes, including cooking oil aerosols, ignition items, organic gaseous pollutants, and steam from the food's moisture content. Emissions from high-temperature frying are classified as probably carcinogenic to humans, and studies indicate an increased risk of lung cancer among cooks [50].

Workers in metal processing, baking, and ship deck crew have been reported to have an increased risk of lung cancer for all histological subtypes; for particular histological subtypes, this risk is higher among construction workers, chefs and cooks, and medical professionals. Odds ratios (ORs) for lung cancer associated with work

as a cook, along with 95% confidence intervals (CIs), were estimated using unconditional logistic regression [51, 52]. The types of lung cancer associated with exposure to cooking oil fumes include adenocarcinoma, squamous cell carcinoma, small cell carcinoma, and large cell carcinoma.

4. Effects of exposure to cooking oil fumes on pregnancy

Exposure to cooking oil fumes during pregnancy significantly increases the risk of having a baby that is large for gestational age. Studies suggest that such exposure can also affect birth weight, with some research indicating a connection between pregnancy exposure to polycyclic aromatic hydrocarbons (PAHs), particulate matter (PM), and volatile organic compounds (VOCs) and abnormal birth weights. Less than 2500 grams at delivery is considered low birth weight and is associated with increased infant morbidity and death that persists throughout adulthood.

Research in California has shown associations between PM_{2.5} exposure and reduced birth weight among full-term infants, even after adjusting for carbon monoxide (CO) levels. Analogously, research conducted in India emphasises the connection between low birth weight and newborn death and mother usage of different domestic cooking fuels. The primary cause of the airborne particulate matter and hazardous chemical levels in traditional open-fire cookstoves is the inefficient burning of coal and biomass. Substantial evidence links household air pollution (HAP) to severe health consequences, especially when exposure occurs in utero or early life [53, 54].

Mounting evidence shows that exposure to HAP during pregnancy is detrimental to maternal health, affects foetal and infant development and viability, and increases the risk of both short- and long-term adverse health and developmental outcomes. Specifically, biomass combustion produces high concentrations of PAHs, which contribute to increased rates of abnormal reproductive system maturation, foetal developmental abnormalities, and growth retardation [54].

Babies born to mothers cooking with wood, dung, or straw were 175 g lighter, on average, compared with babies born to mothers using liquefied petroleum gas (LPG), natural gas, or electricity and hence Household use of high pollution cooking fuels may cause reduced birth weight [55]. According to study findings, children whose moms cooked occasionally, frequently, or consistently during their pregnancies were more likely to exhibit behaviours similar to autism than children whose mothers never cooked. Pregnancy-related cooking frequency and COF exposure raised the likelihood that a kid may exhibit behaviours similar to autism. When mothers used natural gas instead of coal or other cooking fuels, their child's likelihood of exhibiting autistic-like behaviours was reduced. Furthermore, pregnant women using ventilation measures during cooking significantly decreased likelihood of the presence of autistic-like behaviours in their children [56]. In China, with culturally specific cooking habits that are the main source of indoor air pollution the contributions of COFs exposure on preschoolers' neurodevelopmental outcomes may be greater than that found in Western countries [57].

Carbonyl concentrations in kitchens mainly because of its small kitchen and poor ventilation. However, high carbon-containing species (C₅–C₁₀) such as hexaldehyde, heptaldehyde, and nonanaldehyde (16.62–77.00% of total carbonyls) contributed comparatively with low carbon-containing compounds (23.01–83.39% of total carbonyls) in kitchens and the highest total carbonyl concentrations in exhaust streams, indicating that stir-frying and grilling are inclined to produce polluted gases [57].

Exposure to cooking oil fumes (COFs) has been linked to the development of various chronic diseases, including lung cancer, chronic bronchitis, fatty liver, and numerous respiratory symptoms. Additionally, COFs exposure is associated with adverse health outcomes such as oral cancer, cervical intraepithelial neoplasia, chronic bronchitis, and cervical precancerous lesions.

Research has well documented that maternal prenatal exposure to COFs can significantly increase the risk of adverse birth outcomes, including low birth weight (LBW), large for gestational age (LGA), and preterm birth [53, 58–64]. Studies have shown that prenatal exposure to COFs for 0–1 hour per day is significantly associated with LGA newborns, whereas exposure for more than 1 hour per day is associated with small for gestational age (SGA) newborns, although this latter finding did not reach statistical significance [63].

Further studies confirm that prenatal exposure to indoor air pollution (IAP) from biomass, kerosene, or solid fuels can lead to low birth weight, stillbirth, abnormal gestational size, and neonatal death. These findings underscore the global health risks posed by such exposures. Moreover, there is a confirmed correlation between maternal exposure to COFs and children's hyperactive behaviours. Poor ventilation during cooking while mothers were pregnant has been identified as a significant risk factor for clinical levels of hyperactivity in their offspring [65].

Children exhibiting autistic-like behaviours were significantly more likely to have parents with a junior high school education rather than a college degree, mothers who were housewives instead of employed, lower family incomes, and younger maternal and paternal ages [56]. Additionally, as the social economy develops and health awareness improves, there is a gradual shift from polluting to clean cooking fuels. This transition is expected to reduce air pollutants from cooking. Using clean fuels during pregnancy has been shown to reduce the likelihood of hyperactivity in children, compared to using coal fuels [65, 66].

5. Conclusion

Exposure to cooking oil fumes is a significant health concern today. These fumes can generate mutagens and cause cytogenetic changes, making it crucial to select cooking oils that emit fewer harmful substances. Understanding the smoke point of cooking oils is vital, as it influences the generation of harmful compounds. Proper ventilation in cooking environments is essential, as poor ventilation is a major factor in increasing exposure to cooking oil fumes.

In hotels, for example, kitchens often have inadequate ventilation, leading to higher exposure for cooks and bakers, which increases their risk of lung cancer. Pregnant women exposed to cooking oil fumes face risks such as altered gestation periods and having children with conditions like autism, developmental issues, hyperactive behaviour, low birth weight, and preterm delivery.

Economic and educational factors play a significant role in mitigating these risks. As education and economic status improve, people tend to shift towards cleaner and less polluting cooking fuels. However, the fast-food industry, due to its popularity for taste and quick service, exposes workers to higher levels of cooking oil fumes since most fast foods are prepared by frying or deep-frying methods.

Therefore, it is crucial to educate people, particularly less-educated parents-to-be and young fast-food industry workers, about the risks of cooking oil fumes. This education can help reduce lung cancer rates and abnormalities in offspring. Compared to

working mothers, homemakers are more exposed to cooking oil fumes, emphasising the need for awareness and preventive measures in home environments.

Overall, addressing the risks of cooking oil fumes requires a multifaceted approach, including choosing appropriate cooking oils, ensuring proper ventilation, improving economic and educational conditions, and raising awareness about the health impacts of cooking oil fumes. This comprehensive strategy can help mitigate the adverse health effects and improve overall public health.

Author details

Niranjana Karukayil Gopalakrishnan, Sana Syed Ansar, Manikantan Pappuswamy and Aditi Chaudhary*
Department of Life Sciences, Christ University, Bangalore, Karnataka, India

*Address all correspondence to: aditi.chaudhary@res.christuniversity.in

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Lai CH, Jaakkola JJ, Chuang CY, Liou SH, Lung SC, Loh CH, et al. Exposure to cooking oil fumes and oxidative damages: A longitudinal study in Chinese military cooks. *Journal of Exposure Science & Environmental Epidemiology*. 2013;**23**(1):94-100. DOI: 10.1038/jes.2012.87
- [2] Chiang TA, Wu PF, Wang LF, Lee H, Lee CH, Ko YC. Mutagenicity and polycyclic aromatic hydrocarbon content of fumes from heated cooking oils produced in Taiwan. *Mutation Research*. 1997;**381**(2):157-161. DOI: 10.1016/s0027-5107(97)00163-2
- [3] Chiang TA, Wu PF, Ko YC. Prevention of exposure to mutagenic fumes produced by hot cooking oil in Taiwanese kitchens. *Environmental and Molecular Mutagenesis*. 1998;**31**(1):92-96
- [4] Cimrin AH, Alpaydin AO, Ozbal S, Toprak M, Yilmaz O, Uluorman F, et al. Exposure to fumes of a vegetable margarine for frying: Respiratory effects in an experimental model. *ACS Omega*. 2023;**8**(35):31880-31889. DOI: 10.1021/acsomega.3c03340
- [5] Dai X, Lin C, Sun X, Shi Y, Lin Y. The etiology of lung cancer in nonsmoking females in Harbin, China. *Lung Cancer*. 1996;**14**:S85-S91. DOI: 10.1016/s0169-5002(96)90213-5
- [6] Coggon D, Pannett B, Osmond C, Acheson ED. A survey of cancer and occupation in young and middle aged men. I. Cancers of the respiratory tract. *Occupational and Environmental Medicine*. 1986;**43**(5):332-338. DOI: 10.1136/oem.43.5.332
- [7] Kim M, Kim Y, Kim AR, Kwon WJ, Lim S, Kim W, et al. Cooking oil fume exposure and lung-RADS distribution among school cafeteria workers of South Korea. *Annals of Occupational and Environmental Medicine*. 2024;**36**:11-23. DOI: 10.35371/aoem.2024.36.e2
- [8] Wang L, Liu B, Shi L, Yan J, Tan W, Li C, et al. Diverse metabolic effects of cooking oil fume from four edible oils on human BEAS-2B cells: Implications for health guidelines. *Environmental Science & Technology*. 2023;**58**(3):1462-1472. DOI: 10.1021/acs.est.3c05984
- [9] Chen H. A review of the recent research progress on risk factors of lung cancer. *Theoretical and Natural Science*. 2023;**21**(1):291-295. DOI: 10.54254/2753-8818/21/20230905
- [10] Manchanda SC, Passi SJ. Selecting healthy edible oil in the Indian context. *Indian Heart Journal*. 2016;**68**(4):447-449. DOI: 10.1016/j.ihj.2016.05.004
- [11] Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, et al. Dietary fat intake and the risk of coronary heart disease in women. *The New England Journal of Medicine*. 1997;**337**(21):1491-1499. DOI: 10.1056/NEJM199711203372102
- [12] Zhao Y, Wang M, Tao P, Qiu G, Lu X. Emission characteristics of particulate matter from boiling food. *Atmosphere*. 2023;**14**(1):167. DOI: 10.3390/atmos14010167
- [13] Lu F, Shen B, Yuan P, Li S, Sun Y, Mei X. The emission of PM_{2.5} in respiratory zone from Chinese family cooking and its health effect. *The Science of the Total Environment*. 2019;**654**:671-677. DOI: 10.1016/j.scitotenv.2018.10.397
- [14] Ma S, Liu W, Meng C, Dong J, Zhang S. Temperature-dependent

particle mass emission rate during heating of edible oils and their regression models. *Environmental Pollution*. 2023;**323**:121221. DOI: 10.1016/j.envpol.2023.121221

[15] Zhai SR, Albritton D. Airborne particles from cooking oils: Emission test and analysis on chemical and health implications. *Sustainable Cities and Society*. 2020;**52**:101845. DOI: 10.1016/j.scs.2019.101845

[16] Du B, Gao J, Chen J, Stevanovic S, Ristovski Z, Wang L, et al. Particle exposure level and potential health risks of domestic Chinese cooking. *Building and Environment*. 2017;**123**:564-574. DOI: 10.1016/j.buildenv.2017.07.031

[17] Adeniran JA, Yusuf RO, Abdulkadir MO, Yusuf MO, Abdulraheem KA, Adeoye BK, et al. Evaporation rates and pollutants emission from heated cooking oils and influencing factors. *Environmental Pollution*. 2020;**266**:115169. DOI: 10.1016/j.envpol.2020.115169

[18] Kharghani M, Sajjadi S, Moteallehi A, Bargard Z, Naein MA, Kariminezhad F. Investigation of cooking oil quality at fast food restaurants in Mashhad City. *International Journal of Environmental Health Engineering*. 2019;**8**(1):6. DOI: 10.4103/ijehe.ijehe_2_19

[19] Pérez-Palacios T, Petisca C, Casal S, Ferreira IM. Changes in chemical composition of frozen coated fish products during deep-frying. *International Journal of Food Sciences and Nutrition*. 2014;**65**(2):212-218. DOI: 10.3109/09637486.2013.854739

[20] Gertz C. Optimising the baking and frying process using oil-improving agents. *European Journal of Lipid Science and Technology*. 2004;**106**(11):736-745. DOI: 10.1002/ejlt.200401015

[21] Arslan FN, Şapçı AN, Duru F, Kara H. A study on monitoring of frying performance and oxidative stability of cottonseed and palm oil blends in comparison with original oils. *International Journal of Food Properties*. 2016;**20**(3):704-717. DOI: 10.1080/10942912.2016.1177544

[22] Gupta P, Shivhare U, Bawa A. Studies on frying kinetics and quality of French fries. *Drying Technology*. 2000;**18**(1-2):311-321. DOI: 10.1080/07373930008917706

[23] Choe E, Min DB. Chemistry of deep-fat frying oils. *Journal of Food Science*. 2007;**72**(5):R77-R86. DOI: 10.1111/j.1750-3841.2007.00352.x

[24] Anwar F, Hussain AI, Iqbal S, Bhanger MI. Enhancement of the oxidative stability of some vegetable oils by blending with Moringa oleifera oil. *Food Chemistry*. 2007;**103**(4):1181-1191

[25] Ganesan K, Sukalingam K, Xu B. Impact of consumption of repeatedly heated cooking oils on the incidence of various cancers- A critical review. *Critical Reviews in Food Science and Nutrition*. 2017;**59**(3):488-505. DOI: 10.1080/10408398.2017.1379470

[26] Kosmowski M, Pękala-Wojciechowska A, Sut A, Pietras T, Luzak B. Dietary intake of polyphenols or polyunsaturated fatty acids and its relationship with metabolic and inflammatory state in patients with type 2 diabetes mellitus. *Nutrients*. 2022;**14**(5):1083. DOI: 10.3390/nu14051083

[27] Wu PF, Chiang TA, Wang LF, Chang CS, Ko YC. Nitro-polycyclic aromatic hydrocarbon contents of fumes from heated cooking oils and prevention of mutagenicity by catechin. *Mutation Research*. 1998;**403**(1-2):29-34. DOI: 10.1016/s0027-5107(98)00015-3

- [28] Yao Z, Li J, Wu B, Hao X, Yin Y, Jiang X. Characteristics of PAHs from deep-frying and frying cooking fumes. *Environmental Science and Pollution Research International*. 2015;22(20):16110-16120. DOI: 10.1007/s11356-015-4837-4
- [29] Robinson J, Nelson WC. *National Human Activity Pattern Survey Data Base*. Research Triangle Park, NC: USEPA; 1995. p. 2
- [30] Zhao Y, Wang S, Aunan K, Seip HM, Hao J. Air pollution and lung cancer risks in China—A meta-analysis. *Science of the Total Environment*. 2006;366(2-3):500-513. DOI: 10.1016/j.scitotenv.2005.10.010
- [31] Khor YP, Sim BI, Abas F, Lai OM, Wang Y, Wang Y, et al. Quality profile determination of palm olein: Potential markers for the detection of recycled cooking oils. *International Journal of Food Properties*. 2019;22(1):1172-1182. DOI: 10.1080/10942912.2019.1634098
- [32] Alvarenga BR, Xavier FAN, Soares FLF, Carneiro RL. Thermal stability assessment of vegetable oils by Raman spectroscopy and chemometrics. *Food Analytical Methods*. 2018;11(7):1969-1976. DOI: 10.1007/s12161-018-1160-y
- [33] Yen G, Shao C, Chen C, Duh P. Effects of antioxidant and cholesterol on smoke point of oils. *Lebensmittel-Wissenschaft + Technologie/Food Science & Technology*. 1997;30(7):648-652. DOI: 10.1006/fstl.1996.0236
- [34] Woo M, Han S, Song YO. Sesame oil attenuates renal oxidative stress induced by a high fat diet. *Preventive Nutrition and Food Science*. 2019;24(2):114-120. DOI: 10.3746/pnf.2019.24.2.114
- [35] Gunstone F. *Vegetable Oils in Food Technology: Composition, Properties and Uses*. Blackwell Publishing; 2011
- [36] Baseline of Health. 2012. Available from: <https://jonbarron.org/> [Accessed: 26 December 2019]
- [37] What Is the “Truth” about Canola Oil? n.d. Available from: <https://web.archive.org/web/20140413143333/http://www.spectrumorganics.com/shared/faq.php?fqid=34>
- [38] Brenda. Cooking oil and smoking points, whats cooking America. In: *What's Cooking America*. 2023. Available from: <https://whatscookingamerica.net/information/cookingoiltypes.htm>
- [39] Detwiler SB, Markley KS. Smoke, flash, and fire points of soybean and other vegetable oils. *Journal of the American Oil Chemists' Society*. 1940;17(2):39-40. DOI: 10.1007/bf02543003
- [40] Lenz L, Bensch G, Chartier R, Kane M, Ankel-Peters J, Jeuland M. Releasing the killer from the kitchen? Ventilation and air pollution from biomass cooking. *Development Engineering*. 2023;8:100108. DOI: 10.1016/j.deveng.2023.100108
- [41] Iea W, Irena U. WB, Tracking SDG 7: The energy progress report. *The Energy Progress 437 Report 2019*, p. 438. 2019
- [42] Kim C, Gao YT, Xiang YB, Barone-Adesi F, Zhang Y, Hosgood HD, et al. Home kitchen ventilation, cooking fuels, and lung cancer risk in a prospective cohort of never smoking women in Shanghai, China. *International Journal of Cancer*. 2015;136(3):632-638. DOI: 10.1002/ijc.29020
- [43] World Health Organization. *Global Health Observatory (GHO) Data*. WHO Press; 2015
- [44] Centers for Disease Control and Prevention (CDC). *Tobacco use--United*

States, 1900-1999. MMWR. Morbidity and Mortality Weekly Report. 1999;**48**(43):986-993

[45] Epplein M, Schwartz SM, Potter JD, Weiss NS. Smoking-adjusted lung cancer incidence among Asian-Americans (United States). *Cancer Causes & Control* : CCC. 2005;**16**(9):1085-1090. DOI: 10.1007/s10552-005-0330-6

[46] Chen T, Fang Y, Chen H, Chang C, Huang H, Chen Y, et al. Impact of cooking oil fume exposure and fume extractor use on lung cancer risk in non-smoking Han Chinese women. *Scientific Reports*. 2020;**10**(1):1-10. DOI: 10.1038/s41598-020-63656-7

[47] Clément-Duchêne C, Wakelee H. Lung cancer incidence in never smokers. *European Journal of Clinical & Medical Oncology*. 2010;**2**(2):49-57

[48] Yin Z, Cui Z, Ren Y, Zhang H, Yan Y, Zhao Y, et al. Genetic polymorphisms of TERT and CLPTM1L, cooking oil fume exposure, and risk of lung cancer: A case-control study in a Chinese non-smoking female population. *Medical Oncology* (Northwood, London, England). 2014;**31**(8):114. DOI: 10.1007/s12032-014-0114-5

[49] Scarselli A, Marinaccio A, Nesti M. Lavoratori esposti al rischio di tumore del polmone. Una stima a partire dall'archivio ISPEL delle imprese [Exposed workers to lung cancer risk. An estimation using the ISPEL database of enterprises]. *Epidemiologia e Prevenzione*. 2007;**31**(4):212-217

[50] Bigert C, Gustavsson P, Straif K, Pesch B, Brüning T, Kendzia B, et al. Lung cancer risk among cooks when accounting for tobacco smoking: A pooled analysis of case-control studies from Europe, Canada, New Zealand,

and China. *Journal of Occupational and Environmental Medicine*. 2015;**57**(2):202-209. DOI: 10.1097/JOM.0000000000000337

[51] Yenugadhati N, Birkett NJ, Momoli F, Krewski D. Occupations and lung cancer: A population-based case-control study in British Columbia. *Journal of Toxicology and Environmental Health*. Part A. 2009;**72**(10):658-675. DOI: 10.1080/15287390802476892

[52] Ahrens W, Merletti F. A standard tool for the analysis of occupational lung cancer in epidemiologic studies. *International Journal of Occupational and Environmental Health*. 1998;**4**(4):236-240. DOI: 10.1179/oeh.1998.4.4.236

[53] Wang L, Hu W, Guan Q, Du G, Chen T, Wu W, et al. The association between cooking oil fume exposure during pregnancy and birth weight: A prospective mother-child cohort study. *Science of the Total Environment*. 2018;**612**:822-830. DOI: 10.1016/j.scitotenv.2017.08.031

[54] Epstein M, Bates M, Arora N, Balakrishnan K, Jack D, Smith K. Household fuels, low birth weight, and neonatal death in India: The separate impacts of biomass, kerosene, and coal. *International Journal of Hygiene and Environmental Health*. 2013;**216**(5):523-532. DOI: 10.1016/j.ijheh.2012.12.006

[55] Mishra V. Maternal exposure to biomass smoke and reduced birth weight in Zimbabwe*1. *Annals of Epidemiology*. 2004;**14**(10):740-747. DOI: 10.1016/j.annepidem.2004.01.009

[56] Yang J, Strodl E, Wu C, Hou X, Yin X, Wen G, et al. Maternal exposure to cooking oil fumes during pregnancy and autistic-like behaviors in Chinese

- preschoolers. *Environmental Science and Pollution Research International*. 2022;**29**(49):74688-74698. DOI: 10.1007/s11356-022-21113-0
- [57] Yu K, Yang KR, Chen YC, Gong JY, Chen YP, Shih H, et al. Indoor air pollution from gas cooking in five Taiwanese families. *Building and Environment*. 2015;**93**:258-266. DOI: 10.1016/j.buildenv.2015.06.024
- [58] Cheng J, Lee Y, Chen K. Carbonyl compounds in dining areas, kitchens and exhaust streams in restaurants with varying cooking methods in Kaohsiung, Taiwan. *Journal of Environmental Sciences*. 2016;**41**:218-226. DOI: 10.1016/j.jes.2015.06.006
- [59] Lee T, Gany F. Cooking oil fumes and lung cancer: A review of the literature in the context of the U.S. population. *Journal of Immigrant and Minority Health*. 2012;**15**(3):646-652. DOI: 10.1007/s10903-012-9651-1
- [60] Chen C, Zhao Y, Zhao B. Emission rates of multiple air pollutants generated from Chinese residential cooking. *Environmental Science & Technology*. 2018;**52**(3):1081-1087. DOI: 10.1021/acs.est.7b05600
- [61] Lin J, Ni S, Shi Q, Xiong Z, Kang J, Sun X, et al. Environmental exposure to cooking oil fumes and fatty liver disease. *Annals of Palliative Medicine*. 2020;**9**(6):3810-3817. DOI: 10.21037/apm-20-1730
- [62] Hu P, Fan L, Ding P, He Y, Xie C, Niu Z, et al. Association between prenatal exposure to cooking oil fumes and full-term low birth weight is mediated by placental weight. *Environmental Research*. 2018;**167**:622-631. DOI: 10.1016/j.envres.2018.08.014
- [63] Jiang M, Qiu J, Zhou M, He X, Cui H, Lerro C, et al. Exposure to cooking fuels and birth weight in Lanzhou, China: A birth cohort study. *BMC Public Health*. 2015;**15**(1):1-10. DOI: 10.1186/s12889-015-2038-1
- [64] Morales E, Julvez J, Torrent M, De Cid R, Guxens M, Bustamante M, et al. Association of early-life exposure to household gas appliances and indoor nitrogen dioxide with cognition and attention behavior in preschoolers. *American Journal of Epidemiology*. 2009;**169**(11):1327-1336. DOI: 10.1093/aje/kwp067
- [65] Fang X, Strodl E, Wu C, Liu L, Yin X, Wen G, et al. Maternal cooking during pregnancy may increase hyperactive behaviors among children aged at around 3 years old. *Indoor Air*. 2019;**30**(1):126-136. DOI: 10.1111/ina.12614
- [66] Shupler M, Hystad P, Gustafson P, Rangarajan S, Mushtaha M, Jayachtria KG, et al. Household, community, sub-national and country-level predictors of primary cooking fuel switching in nine countries from the PURE study. *Environmental Research Letters*. 2019;**14**(8):085006. DOI: 10.1088/1748-9326/ab2d46

β -Thalassemia in India: A Classic Case of Genetic Adaptation in Response to a Pathogen with a Historical Perspective on Ancient Population Migrations (a Mini-Review)

Chanda Siddoo-Atwal

Abstract

In the 1940s, JBS Haldane hypothesized that haemoglobinopathies including β -thalassemia had become common in regions where malaria was endemic and that natural selection had acted to increase their prevalence. Subsequent case-control studies have confirmed that both α - and β -thalassemia provide a high degree of protection against clinical malaria. Then, in 1956, a hematologist, JK Siddoo, published a pioneering paper reporting that β -thalassemia (also known as Mediterranean anemia as it is common among Greeks, Italians, and other Mediterranean peoples) was prevalent in Sikh populations from the Punjab. The hypothesis was that invading Aryan tribes had brought the gene to India with them around 1500 BC from the southern steppes of Europe. However, there is little evidence to support this theory. Nevertheless, nomadic Iranian (Eurasian) tribes known as “Scythians” started to invade the subcontinent in succession around 50 BC, following the incursions of Alexander the Great in 325 BC. Moreover, the crumbling Indo-Greek kingdom that had persisted for almost 200 years finally fell to these invaders from the East and some of the displaced Greeks intermarried with the conquering Scythians. Therefore, sequencing and comparative analysis of the beta-globin gene from modern-day populations could help to provide conclusive proof for this ancient instance of gene flow and genetic adaptation to a dangerous pathogen.

Keywords: α -/ β -thalassemia, gene regulation, single base substitutions, frameshift mutations, RNA-processing mutants, transcriptional mutants, initiation codon mutations

1. Introduction

Historically, most of its invaders have entered India through the northwest corridor of the subcontinent. It has been proposed that the first invaders of Punjab in India were the Aryan tribes between 2000 and 1500 BC (sharing a common place of origin and pre-history with the Iranians). However, there is virtually no archaeological evidence to support this claim, which is only substantiated through the new Indo-European language the invaders brought with them to the subcontinent, (Sanskrit) and the existence of sacred writings and rituals still performed in this language to this day. There are also some intriguing genetic studies based on comparative analysis of the Y chromosome that suggest populations with the Y-haplogroup R1a1 moved to the Russian Plain around 4800–4600 B.C from where they migrated to Anatolia, Mittani, the Arabian Peninsula, Iran, the South Urals, and North India [1].

The primitive culture of these Aryans was grafted onto the highly sophisticated civilization of the Indus Valley, and, in time, they developed a caste system to separate the Aryan conquerors from the conquered Dravidians. The former was divided into Brahmins (priests), Kshatriyas (warriors), and Vaisyas (traders and farmers), while the latter was restricted to the Shudra or “untouchable” classes. Ultimately, the Brahmins, by virtue of their wit and learning, came to rule supreme over all the other castes or social groups, which flexible at first, eventually came to be determined by birth [2, 3].

The first evidence of a historical invasion occurs c. 500 BC from the writings of Greek historians such as Herodotus and Strabo who mention a Persian kingdom in northern India. It was Darius the Great who added northwestern India to the Achaemenid Empire during his reign (522–486 BC) [4].

Following the Indian exploits of Alexander the Great in 325 B.C. that have been very well-documented by Greek historians such as Arrian, various Greek satraps and kings governed over the Punjab region for a period of 150–200 years. They are commemorated by a wealth of coins and in Sanskrit literary works like the Ramayana in which they are called “Yavanarajas.” Perhaps, the most notable of these monarchs was Menander, who became a Buddhist, and ruled over a larger kingdom than any other known Indo-Greek rulers as documented by the Greek biographer, Plutarch [2, 3, 5].

After the Greeks, various nomadic Iranian-speaking Scythian tribes from Central Asia were the next set of conquerors to enter the portals of northwest India between 50 B.C. and 150 A.D. For the most part, the legacy of the Greeks was not a long-lasting one and traces of their culture were eroded with the advent of the mass immigration of these Scythian hordes and their subsequent inter-marriages with the Greeks. However, in the Kanishka era of the Kushanas, the relatively late (easternmost) Scythian tribe to enter India with distinctly Turkic cultural elements, the Gandhara style of sculpture, which was heavily influenced by the Greek school, flourished in northwest India [6, 7]. Many of the coins minted by the earlier Scythian, Parthian, and Kushana kings had also featured Greek deities such as Apollo, Poseidon, and Zeus suggesting a shared theism and some degree of cross-cultural assimilation. As these rulers all belonged to different Scythian tribes, this part of Indian history is known as *The Scythian Period* [7].

In the annals of one of the leading Rajput families of Rajasthan (Bhatti), a branch of which was supplanted to Punjab as one of its royal families (Sidhu-Brar), one can start to distinguish traces of their Scythian origin. A fourth century AD inscription mentioned in the Bhatti history (Annals of Jessulmer) states that their ancestor Saca Salivahan, a “Gete” or Scythic prince ruled over Salpoor and confirms this assertion

of a Scythian descent [8]. In time, as many of the Rajput clans such as the Bhattis were transplanted to the more fertile region of the Punjab, they assumed the name of “Jat,” the appellation of the local farming communities, which also claimed descent from Gete or Scythic invaders [8–11]. Genetic studies suggest a more recent relationship to Europe and Central Asia of certain “Jat” populations due to a higher complement of shared DNA and, thus, support a later historic migration of some groups to the subcontinent [11, 12]. Similarly, genetic studies on ancient mummies from the Tarim Basin in the region that was once inhabited by the Yuezhi and the Kushanas have revealed a Eurasian genetic profile [13–15]; even today, the main modern Turkic populations in this Tarim basin region of Xinjiang are a genetic mixture of European and East Asian lineages found in contemporary Central Asians [15, 16]. In addition, a significant overlap of the DNA of Jats with that of earlier North Indian “Indo-European” groups also supports the Russian steppe hypothesis and a prehistoric Aryan migration into the Indus Valley as they were all related Indo-Aryan/Indo-Iranian tribes with a shared theism [12].

Starting in the second millennium, invaders of Punjab and northwest India included the Turks, Pathans (Afghans), and Mughals (Persianized descendants of the Mongols), who brought with them a highly elaborate Islamic culture of their own. Not only did they resist Hinduization until the later reign of the “Great Mughals” (particularly, Akbar, who married a Rajput princess and adopted Indian customs) but also they often cruelly tried to force Islam on the Indian masses as their holy mission [17]. This led to the rise of various religious reform movements in medieval India, such as Sikhism, which advocated spiritual universality and social equality. Ironically, Guru Nanak had started a socialistic movement (in which women were regarded as equals) in reaction to the tyranny of both the Brahmins and the Islamic invaders. It would become most popular with the “Jat” farming communities of Punjab who already had a more egalitarian social structure than other parts of India [9, 18].

It is in this milieu that the Sikhs emerged as a significant political power in Northwest India in the eighteenth century (prior to the British in the mid-nineteenth century) culminating with the empire of Maharaja Ranjit Singh, who actively employed women in his administration and annexed parts of Afghanistan [19, 20].

2. Introduction to β-thalassemia and its history

β-Thalassemias are a group of hereditary blood disorders affecting the synthesis of β-hemoglobin chains, and the results range from severe anemias requiring regular blood transfusions to asymptomatic individuals. There are three main forms of the disease: thalassemia major, the most severe form with early presentation; thalassemia intermedia, a moderate form with clinical symptoms that present later in life; and thalassemia minor, the least severe and asymptomatic form, which can cause moderate anemia. Frequency of this disease is 1 in 10,000 in the European Union and 1 in 100,000 throughout the rest of the world [21].

In developing countries where diagnosis of thalassemia major can be delayed, clinically, children may display growth retardation, pallor, jaundice, poor musculature, leg ulcers, and skeletal changes among other symptoms. In developed countries where blood transfusions may be implemented sooner, patients can develop iron overload-related issues including endocrine complications, dilated cardiomyopathy, liver fibrosis, and cirrhosis. The main clinical features of thalassemia intermedia patients are hypertrophy of erythroid marrow and related complications, gallstones,

painful leg ulcers, and an increased predisposition to thrombosis. Thalassemia minor is not associated with any clinical symptoms except mild or moderate anemia in certain cases [21].

The effects of thalassemia major on the body are multi-faceted and extremely complex. Treatment involves regular blood transfusions administered every 2 to 5 weeks in order to promote normal growth in children, normal physical activity, adequately suppress bone marrow activity, and minimize transfusional iron accumulation. Complications can include increased transfusion requirements, fever due to a variety of bacterial infections including *Yersinia* and *Klebsiella*, back pain due to osteoporosis, unexplained abdominal pain and chest pain due to a variety of factors, dyspnoea, jaundice due to liver failure, and leg cramps due to hypoparathyroidism [22] and most probably due to a lack of oxygen in the muscles as experienced by athletes.

India has a large β -thalassemia burden, and there are an estimated 100,000 patients with a syndrome, but in the absence of National Registries of patients, the exact numbers are unknown. Recent studies estimate the average prevalence of β -thalassemia carriers at 3–4% [23], while an extensive screening project based on a large number of college students and pregnant women in the states of Maharashtra, Gujarat, Punjab, Karnataka, West Bengal, and Assam indicated a national average carrier rate of 2.78% [24]. Nation-wide screening programs in different regions have revealed that many ethnic groups such as Sindhis, Kutchi Bhanushalis, Lohanas, Punjabi Khatri and Vaisya-Aroras (it is quite possible that these Indo-Aryans of the Punjab intermarried with the Greeks in pre-Scythian times), Bengalis, certain Muslim groups, and tribal populations from Gujarat and Orissa have a higher prevalence ranging from 4 to 17%. It is suggested by the authors of a recent report that there is a need for education and awareness programs for health professionals, college students, pregnant women and the general public, a requirement to intensify screening in all states and to provide genetic counseling and prenatal diagnosis [23]. The main treatments that are available include blood transfusions and iron chelation therapy. While India bears approximately 25% of the global β -thalassemia burden, the Punjab with an estimated 4700 thalassemia major patients contributes nearly 5% of the national total [25, 26]. As a result, Project Rainbow has been initiated there to try and reduce this burden. Affected individuals often display complications like growth retardation, endocrine and cardiac dysfunction, and transfusion-transmitted infections that arise as a result of iron overload and the use of unsafe blood [26].

2.1 Old World origin of the gene

It is unlikely that humans were exposed to large malaria outbreaks before the advent of agriculture when early Neolithic farmers settled in mosquito-infested soft and marshy soil near standing water [27]. Recent studies from ancient sites of the Middle Euphrates region support the view that they are contemporary with the beginning of cultivation (c. 10,000 B.C.) and that the world's first urban centers developed in the late fourth millennium B.C. along the banks of the Tigris and Euphrates Rivers in Mesopotamia. This region includes southern Iraq, a small region of southeast Turkey and northeast Syria around the Middle Euphrates. This may also be the birthplace of wheat agricultural innovation [27, 28].

After the domestication of wheat, the agricultural technology spread west through the Mediterranean basin. As agriculture spread in Greece, it is possible that Greece was one of the most malarious regions of the Mediterranean. Subsequently, Anatolia was the site of several civilizations including Hittite, Persian, Greek, Roman,

Byzantine, and Seljuk between 2000 B.C. and 1100 A.D. resulting in the flow of various mutations into and out of present-day Turkey, which became cross-roads for the old world [27].

Thus, the most common and widespread β -thalassemia mutations in the Mediterranean World and adjoining Arab countries are likely the oldest. For example, Codon 39 (C > T) is believed to be of Roman origin, and IVS-1-110 is believed to have arisen in Turkey. In fact, malarial selection for this oldest β -thalassemia allele in Anatolia may date to a period somewhere between 6500 and 2000 BC. The latter reaches its highest frequency in Eastern Mediterranean Arab countries and may have been introduced there by Turkish, Greek, or Phoenician settlers [29]. Interestingly, the β -thalassemia mutations reported in Arab populations are mainly of Mediterranean and Asian origin, suggesting ancient commerce and trade routes [27].

At the same time, the malaria antigen has also been detected in Egyptian mummies dating from 3200 and 1304 BC suggesting another possible site of origin [30]. According to this theory, malaria most likely traveled from the African rainforest down the Nile to the Mediterranean spreading east to the Fertile Crescent and north to Greece. Greek traders and colonists brought it to Italy from where Roman soldiers and merchants would carry it as far north as England and Denmark [31]. However, β -thalassemia mutations are relatively rare in Africa [32] while another inherited hemoglobin disorder, sickle cell anemia, is much more common [33]. Therefore, the African continent is more likely the place of origin for sickle cell mutations, especially, as consanguineous marriages are popular there.

3. DNA mutations and haplotypes associated with gene flow and migration

An intriguing facet of modern genetics is that genetic polymorphisms that tend to be inherited together (DNA haplotypes) can help to determine the date of origin, track gene flow of a particular β -globin gene mutation, and detect major migrations [27]. Unfortunately, extensive studies in this area are lacking on the Indian subcontinent.

In general, it is estimated that 1.5% of the global population are carriers of β -thalassemia. The average frequency of β -thalassemia carriers in Greece is 7.4%, but in some regions, it can be as high as 15–20%. Interestingly, in the small Sikh study carried out by JK Siddoo in the 1950s, the mean frequency was at least 6% corresponding quite closely to the levels observed in Greek communities [34, 35]. Although there do not appear to be too many contemporary ethnic studies on this subject in India, there is one large Indian study that corroborates this earlier finding that the prevalence of β -thalassemia in the country is highest in Punjabi and Sindhi (originating from the adjoining province of Sindh in Pakistan) communities [36, 37]. Although the major Greek mutations have been well documented, the sequencing data from India is not as comprehensive, and large-scale studies in specific states of interest are warranted [38]. Nevertheless, one interesting observation to arise is that the people of Greece and Turkey have had a shared genetic lineage since prehistory (Neolithic times) [39]; a shared history since the time of Troy; and Greeks have colonized Turkey from the time of Alexander the Great. Thus, one common promoter region mutation has been identified among Jat-Sikhs (46%) that coincides with a Turkish genotype, –88 (C-T) [40, 41]. In Turkey, there is also a relatively high prevalence of the β -thalassemia trait in some areas (up to 7.9%). A further Indo-Greek collaboration could help to shed

more light on this ancient mystery and uncover other possible Mediterranean mutations that are common between these two populations.

Studies from Pakistani populations claiming Greek descent are a little more revealing [42]. Using genetic distance measures sensitive to recent events, the lowest distances were observed between Pathans and Greeks. Clade E3b1 lineages which are frequent in Greeks, but not in Pakistanis, were observed in two Pathan individuals, one of whom shared a 16 Y-STR haplotype with Greeks. A shortened version of this haplotype is concentrated in Macedonia and Greece providing strong evidence for a European origin for at least a small percentage of the Pathan Y chromosomes. A close genetic affinity has also been found between Jat and Afghani populations, particularly Pathans, since they share a joint culture and history [12]. In the past, Punjab has been a province of Afghanistan; parts of Afghanistan have been annexed by Punjab, and the Sikhs regularly fought territorial wars with the Pathans [19]. So, not only were the borders between the two areas fluid but also most invaders entered India through the Northwest and several Jat tribes are known to have inhabited Afghanistan [11, 43]. Thus, a genetic overlap in these two groups is not surprising.

So far, in India, it has been established through genetic studies that there are five common mutations, which account for between 80 and 93% of all β -thalassemia mutations, and five less common mutations including the “Turkish” mutation prevalent among Jat-Sikhs. The results suggest relatively recent and local origins for most β -thalassemia mutations. In addition, multiple haplotype associations and widespread geographical distribution of certain high frequency alleles are also consistent with an older origin. In this regard, another very interesting case of possible gene flow in the opposite direction involving β -thalassemia mutations from the Indian subcontinent occurs in the Omani population. The nature of the β -thalassaemia mutations that have been determined in the Omani population suggests that the majority have been introduced by gene flow. There was active commerce between Oman and the area around the Indus Valley for centuries. Burial sites and pottery dating to around 3000 B.C. have also been unearthed in the surrounding Gulf countries. Using archaeological data from all these various sites, it has been possible to establish that there was trade between Mesopotamia (present day Iraq), the lands bordering the Arabian Gulf and the Indus Valley people. Historically, Oman was a major maritime nation with links to the Far East and India on the one hand and East Africa and Egypt on the other, as well as trade in the Arabian Gulf. IVS-1-5 (G > C), a severe β^+ allele, is the most prevalent of the mutations thus far described in Oman. A detailed analysis of the distribution of β -thalassemia mutations in Pakistan shows that the IVS I-5 (G > C) mutation is widespread in that country, reaching a peak incidence in the Baluchis (76.2%) and having a high prevalence among the Sindhis (43.9%). It is also found at a high frequency in India [27, 44]. Thus, it is highly likely that altogether, the β -thalassemia mutations on the Indian subcontinent reflect considerable ethnic diversity, and gene flow from population migrations particularly on the Northwest route of invaders, and natural selection by malaria infection [45, 46].

4. A brief analysis of β -thalassemia and β -globin gene mutations

β -Thalassemia is represented by the decreased production of β -globin chains in the adult hemoglobin A molecule, which is involved in oxygen delivery to the entire body. The result of this biochemical anomaly is inefficient erythropoiesis and peripheral hemolysis [27]. Physiologically, an inefficient circulatory system and inability of

blood to carry oxygen efficiently to every cell and organ is tantamount to a long-term death sentence.

The main cause of β-thalassemias is point mutations and, occasionally, deletions of the β-globin gene on chromosome 11. This genetic disorder is predominantly inherited in an autosomal recessive manner, although, more rarely, there are also dominant mutations of the gene. β-thalassemia is particularly prevalent in regions where malaria is endemic and confers an adaptive advantage against infection [47, 48].

More than 180 mutations have been identified and classified in the complex β-globin gene, which requires tight control in the synthesis of its tetrameric hemoglobin molecules [49–51]. This gene contains three exons and two introns and displays several levels of transcriptional and translational regulation. β-Thalassemia mutations include nonsense mutations, RNA processing mutations, transcriptional mutations in the promoter region, and mutations in the initiation codon [52–54].

This is a vast field of study, and much research has been conducted in it. In fact, the well-elucidated structure of the β-globin gene including its promoter, enhancer, start and stop codons, translated and untranslated regions, GATA sites for transcription factors and poly(A) site serves as an excellent model for genetic studies in general [55].

As an example of how migration detection in β-thalassemia carriers works, a study in Germany with non-immigrant German populations found that roughly two-thirds of thalassemia minor cases appear to be caused by Mediterranean mutations (61%). In contrast, approximately one-third of cases may have originated as a result of local mutations such as the one that affects position –2 of the intron 1 splice acceptor site (IVSI-129 A-G) and a deletion of a single G in codon 15/16 (FS 15/16 ΔG). The former finding suggests introduction due to migration from the Mediterranean region, probably during the historic Roman occupation of northern Europe [41].

Similarly, while common mutations causing β-thalassemia major and minor occur in populations from northern, western, and eastern India, rare and even novel mutations may occur in some of these populations. In one Indian study, DNA sequencing of LCR HS2, 3, and 4 core sequences revealed a polymorphism, an A-G, in the palindromic sequence, TGGGGACCCCA, of LCR HS4 [56]. So, once again, the G allele could be a new evolutionary mutation in the Indian population, while certain other Mediterranean mutations may have been introduced via migration. However, outbreeding is rare in India due to the highly rigid caste system, which does not encourage inter-caste unions. Therefore, one opportunity for this to have occurred is as a result of the intermarriages documented to have taken place during the Indo-Greek and Scythian periods of Indian history in the frontier provinces [57–59]. There is a good possibility of finding overlapping mutations in this select gene pool with the genetic profile of “recent” Eurasian nomads that might be expected from Indo-Scythians [11, 13, 55].

5. Future perspectives

Perpetuating a painful and debilitating genetic disorder like β-thalassemia major that is a slow death sentence does not seem humane. Extensive genetic counseling and prenatal diagnosis campaigns appear to be the best options available to countries with a significant population afflicted by this devastating disease to reduce its burden. It can also constitute a moral dilemma and financial burden for the parents of young patients.

One future implication is that, with the progress of climate change and an increase in air pollution, the condition of β -thalassemia homozygotes will likely worsen as the frequency of blood transfusions will increase with their increasing oxygen needs. At the same time, it is a real possibility that heterozygotes may start to experience some of the same symptoms as homozygotes such as muscle cramps from pollutants in the atmosphere that interfere with oxygen delivery (personal observation). Most heterozygotes, who are not anemic, are often on the lower end of normal blood production, which may no longer be an adequate hemoglobin level for normal physiological functioning with the decreasing air quality in the world. According to the Air Quality Index, there is hardly any country on the Earth with pristine air that is not polluted with coarse (PM10) and fine (PM2.5) particulate matter (from wildfire smoke and fossil fuels linked to lung irritation and respiratory tract cancer), nitrogen dioxide and carbon monoxide (which decrease the capacity of hemoglobin to carry oxygen), and ozone (which can reduce lung function and inflame the lining of the lungs) [22, 60–64]. Another common industrial air pollutant is sulfur dioxide, which irritates the respiratory tract with particularly serious effects in asthmatics [65].

A recent EU report has stated that there were 400,000 deaths in Europe in 2021 related to three main air pollutants and that some of these fatalities could have been avoided if pollutants had been reduced to meet WHO standards. Pollution caused by fine particulate matter (PM2.5) was associated with 253,000 deaths, while 52,000 deaths were attributed to nitrogen dioxide pollution, and 22,000 deaths were attributed to ozone exposure [66].

As a preventive measure, the use of air purifiers indoors and an app for thalassemics and others suffering from blood disorders that calculates the levels of toxic gases in the outside air and when it is safe for them to go outdoors may be helpful in their daily lives.

6. Conclusions

In conclusion, the multiple levels of gene regulation, as illustrated in this example of the β -globin genes, make up a highly elaborate and finely tuned system to determine subtle differences in gene expression between individuals resulting in varied production and type of β -globin chains. Such small differences or genetic polymorphisms are the reason for the genetic susceptibility of certain individuals to specific diseases and may represent an ancient inherited adaptive response to a killer pathogen like the *Plasmodium* parasite in India. It is highly likely that altogether the β -thalassemia mutations on the Indian subcontinent reflect considerable ethnic diversity, gene flow from population migration particularly in the Northwest (probably represented by the less common mutations), and natural selection by malaria infection. Genetic studies suggest relatively recent and local origins for most β -thalassemia mutations, while multiple haplotype associations and widespread geographical distribution of certain high frequency alleles (common mutations) are also consistent with an older origin.

Given the varied history of Punjab, it is not surprising that a “Turkish” β -thalassemia mutation would gain prevalence among the population. Nevertheless, it is also an intriguing possibility that this same gene pool, which likely absorbed the Indo-Greeks, became the repository for an older Mediterranean mutation from Ancient Greece. The limited study in Pathan men, which could be expanded and extended to include Jat-Sikh men, may point to a possible link to Greece and Macedon.

There are also still more β -thalassemia mutations to be identified in the Sikh community that may correspond with Greek ones. Naturally, this is a subject which needs to be actively studied by pursuing a general campaign of gene sequencing and a more detailed study of possible accompanying DNA haplotypes in India with a special emphasis on their link to specific past migrations [45, 46, 67]. In either case, whether it represents an old Mediterranean mutation or a Central Asian one, it is a fascinating finding that links this particular population to its past heritage, especially as gene sequencing of Turkish genomes in Turkey has revealed an affinity with both South European populations (which encompasses Greeks and Mediterraneans) plus a relatively more recent contribution from ancestral East Asian populations [68]. The “Turkish” β -thalassemia mutation also correlates well with the prevalence of Y-chromosome Haplogroup Q common to Huns, Mongols, and Turks among Jats [11].

Furthermore, in the German study identifying the “Turkish” β -thalassemia mutation, two other participants, who were listed as being of unspecified Middle Eastern origin, also appeared to share the same mutation [41]. It does not appear to be a common mutation in Arab populations, though However, Turkey (which is sometimes classified as Mediterranean and sometimes as Middle Eastern) has the highest average prevalence of the β -thalassemia trait at 2.7% than any of the Middle Eastern countries and ranges between 5 and 7.9% on its East Mediterranean seaboard. Therefore, this incidence level is more in keeping with that observed in Sikh populations [6%] and is also in agreement with the recorded history of Punjab, its people, and their combined European (mainly Scandinavian and Balto-Slavic) and Central Asian (mainly Mongolian and Turkic Central Asian republics) genetic profile [11, 27, 37].

Finally, in Iranian populations, the prevalence of β -thalassemia is estimated at approximately 4–8% in most areas, while in the Mazandaran province 10% of the population are carriers; however, the “Turkish” mutation is not one of the common mutations cited in the available literature [69]. Persia was also part of Alexander the Great’s eastern empire, and this might help explain the high incidence of β -thalassemia carriers in Iran where intermarriages between Persians and Greeks were documented to have occurred, as well, by Greek historians. Nevertheless, the genetic profile of Sikhs fits with what one might expect of Scythian (Eurasian) nomads from the far eastern reaches of the Persian Empire, whose domain extended all the way from China to Central Europe according to Strabo and archaeological excavations [70, 71].

Such complicated histories hardly make for facile genetic studies. However, the results can be tantalizing, especially when they support the writings of classical historians and literary figures from the ancient world including Greece, China, Persia, and India.

Acknowledgements

This paper is dedicated to my aunt, Dr. JK Siddoo, the great female role-model in my life. The author would also like to thank the organizers of the 16th International Conference on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases for inviting me to present the Abstract in its current form as part of a poster in Dresden, Germany, at which helpful discussions took place and to acknowledge oral presentation of an altered Abstract with a brief summary of the same data at the 51st Global Congress on Infectious Diseases, SciTech online conference.

Author details

Chanda Siddoo-Atwal^{1,2,3,4}

1 University College London, UK

2 Simon Fraser University, Canada

3 Medical College of Wisconsin, USA

4 Moondust Cosmetics Ltd., Canada

*Address all correspondence to: moondustcosmetics@gmail.com

IntechOpen

© 2024 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Klyosov A, Rozhanskii I. Haplogroup R1a as the proto Indo-Europeans and the legendary Aryans as witnessed by the DNA of their current descendants. *Advances in Anthropology*. 2012;**2**:1-13. DOI: 10.4236/aa.2012.21001
- [2] Dunbar G. *History of India*. London, England: Nicholson and Watson; 1949
- [3] Ghosh NN. *Early History of India*. Allahabad, India: Indian Press; 1951
- [4] Department of Ancient Near Eastern Art. The Achaemenid Persian empire (550-330 B.C.). In: Heilbrunn Timeline of Art History. New York: The Metropolitan Museum of Art, 2000; 2004. Available from: http://www.metmuseum.org/toah/hd/acha/hd_acha.htm [Accessed: September 12, 2024]
- [5] Galiana J. The Greek Rajas of Ancient India. *Hellenistic History*; 2022. Available from: www.hellenistichistory.com/2022/09/24/the-greek-rajas-of-ancient-india/ [Accessed: September 12, 2024]
- [6] The Sui Dynasty. Available from: <https://depts.washington.edu/silkroad/exhibit/sui/essay.html> [Accessed: September 12, 2024]
- [7] van Lohuizen-de Leeuw JE. *The Scythian Period*. Leiden, Netherlands: Leiden; 1949
- [8] Tod J. *Annals and Antiquities of Rajasthan*. Oxford, England: Oxford University Press; 1920
- [9] Griffin LH. *The Rajas of the Punjab*. London: Trubner & Co.; 1873
- [10] Johnston HJM. *Jewels of the Qila*. West Mall, Vancouver, Canada: UBC Press; 2012
- [11] Mahal DG, Matsoukas IG. Y-STR Haplogroup diversity in the Jat population reveals several different ancient origins. *Frontiers in Genetics*. 2017;**20**(8):121. DOI: 10.3389/fgene.2017.00121
- [12] Pathak AK, Kadian A, Kushniarevich A, et al. The genetic ancestry of modern Indus Valley populations from Northwest India. *The American Journal of Human Genetics*. 2018;**103**(6):918-929. DOI: 10.1016/j.ajhg.2018.10.022. ISSN 0002-9297
- [13] Wei R, Li H, Wenkan X. The separate origins of the Tocharians and the Yuezhi: Results from recent advances in archaeology and genetics. In: Malzahn M, Peyrot M, Fellner HA, Illés T-S, editors. *Tocharian Texts in Context International Conference on Tocharian Manuscripts and Silk Road Culture*, June 26-28. Hempen; Jan 2015. ISBN: 978-3-944312-26-2
- [14] Li C, Li H, Cui Y, Xie C, Cai D, Li W, et al. Evidence that a west-east admixed population lived in the Tarim Basin as early as the early bronze age. *BMC Biology*. 2010;**17**(8):15. DOI: 10.1186/1741-7007-8-15
- [15] Li C, Ning C, Hagelberg E, Li H, Zhao Y, Li W, et al. Analysis of ancient human mitochondrial DNA from the Xiaohe cemetery: Insights into prehistoric population movements in the Tarim Basin, China. *BMC Genetics*. 2015;**8**(16):78. DOI: 10.1186/s12863-015-0237-5
- [16] Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, et al. Admixture, migrations, and dispersals in Central Asia: Evidence from maternal DNA lineages. *European Journal of Human Genetics*. 2004;**12**(6):495-504. DOI: 10.1038/sj.ejhg.5201160

- [17] Grewal JS. The Sikhs of the Punjab. Cambridge, UK: Cambridge University Press; 1991
- [18] Sher SS. Sikhism and the Sikhs. In: Dunne FP, editor. The World Religions Speak on the Relevance of Religion in the Modern World. Dordrecht: World Academy of Art and Science. Springer; 1970. DOI: 10.1007/978-94-017-5892-5_14
- [19] Singh K. Ranjit Singh: Maharajah of the Punjab. Gurgaon, Delhi, India: Penguin India; 1963
- [20] Gupta HR. History of the Sikhs. S.N. Sarkar; 1944
- [21] Galanello R, Origa R. Beta-thalassemia. Orphanet Journal of Rare Diseases. 2010;5:11. DOI: 10.1186/1750-1172-5-11
- [22] Cappellini MD, Cohen A, Eleftheriou A, et al. Chapter 18. In: Guidelines for the Clinical Management of Thalassaemia [Internet]. 2nd Revised ed. Nicosia (CY): Thalassaemia International Federation; 2008; Outline of Diagnostic Dilemmas in Thalassaemia
- [23] Colah R, Italia K, Gorakshakar A. Burden of thalassemia in India: The road map for control. Pediatric Hematology Oncology Journal. 2017;2:79-84. DOI: 10.1016/j.phoj.2017.10.002
- [24] Mohanty D, Colah R, Gorakshakar A, Vigyan J, A National Multicentric Task Force Study of ICMR (2000-2005). S & T Mission Project on Community Control of Thalassaemia Syndromes—Awareness, Screening, Genetic Counselling and Prevention. New Delhi: Indian Council of Medical Research; 2008
- [25] Singh P, Shaikh S, Parmar S, Gupta R. Current status of β -thalassemic burden in India. Hemoglobin. 2023;47(5):181-190. DOI: 10.1080/03630269.2023.2269837. Epub 2023 Dec 7
- [26] Mathew A, Sobti PC. The burden of thalassemia in Punjab: A roadmap forward. Pediatric Hematology Oncology Journal. 2017;2(4):85-87. DOI: 10.1016/j.phoj.2018.01.001, ISSN 2468-1245
- [27] De Sanctis V, Kattamis C, Canatan D, Soliman AT, Elsedfy H, Karimi M, et al. β -Thalassemia distribution in the old world: An ancient disease seen from a historical standpoint. Mediterr J Hematol Infect Dis. 2017;9(1):e2017018. DOI: 10.4084/MJHID.2017.018
- [28] Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R, Martin W. Genetics and geography of wild cereal domestication in the near east. Nature Reviews. Genetics. 2002;3:429-441
- [29] Zahed L. The spectrum of beta-thalassemia mutations in the Arab populations. Journal of Biomedicine & Biotechnology. 2001;1:129-132. DOI: 10.1155/S1110724301000298
- [30] Miller RL, Ikram S, Armelagos GJ, Walker R, Harer WB, Shiff CJ, et al. Diagnosis of *plasmidium falciparum* infections in mummies using the rapid manual ParaSight-F test. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1994;88(1):31-32
- [31] Karlen A. Man and Microbes: Disease and Plagues in History and Modern Times. New York: G.P. Putnam; 1995
- [32] Rao E, Chandraker SK, Singh MM, Kumar R. Global distribution of β -thalassemia mutations: An update. Gene. 2024;896:148022. DOI: 10.1016/j.gene.2023.148022. ISSN 0378-1119
- [33] Institute of Medicine (US) Committee on the Economics of Antimalarial Drugs. In: Arrow KJ,

Panosian C, Gelband H, editors. Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance. Washington (DC): National Academies Press (US); 2004. 5, A Brief History of Malaria. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK215638/>

[34] Loukopoulos D. Haemoglobinopathies in Greece: Prevention programme over the past 35 years. The Indian Journal of Medical Research. 2011;**134**(4):572-576

[35] Siddoo JK, Siddoo SK, Chase WH, Morgan-Dean L, Perry WH. Thalassemia in Sikhs. Blood, The Journal of Hematology. 1956;**XI**(3):197-209

[36] Sinha S, Black ML, Agarwal S, Colah R, Das R, Ryan K, et al. Profiling β-thalassaemia mutations in India at state and regional levels: Implications for genetic education, screening and counselling programmes. The HUGO Journal. 2009;**3**(1-4):51-62. DOI: 10.1007/s11568-010-9132-3

[37] Tozun M, Turhan E, Babaoglu AB. Beta thalassemia trait in Turkey and the Middle East: A meta-analysis of prevalence. Acta Medica Mediterranea. 2018;**34**:1731-1738

[38] Georgiou I, Makis A, Chaidos A, Bouba I, Hatzi E, Kranas V, et al. Distribution and frequency of beta-thalassemia mutations in northwestern and Central Greece. European Journal of Haematology. 2003;**70**(2):75-78. DOI: 10.1034/j.1600-0609.2003.00017.x

[39] Hofmanova Z, Kreutzer S, Hellenthal G. Early farmers from across Europe directly descended from Neolithic Aegeans. Anthropology. 2016;**113**(25):6886-6891. DOI: 10.1073/pnas.1523951113

[40] Garewal G, Das R, Ahluwalia J, Marwaha RK, Varma S. Nucleotide –88

(C-T) promoter mutation is a common beta-thalassemia mutation in the Jat Sikhs of Punjab, India. American Journal of Hematology. 2005;**79**(4):252-256. DOI: 10.1002/ajh.20445

[41] Vetter B, Schwarz C, Kohne E, Kulozik AE. Beta-thalassaemia in the immigrant and non-immigrant German populations. British Journal of Haematology. 1997;**97**(2):266-272

[42] Firasat S, Khaliq S, Mohyuddin A, Papaioannou M, Tyler-Smith C, Underhill PA, et al. Y-chromosomal evidence for a limited Greek contribution to the Pathan population of Pakistan. European Journal of Human Genetics. 2007;**15**(1):121-126. DOI: 10.1038/sj.ejhg.5201726. Epub 2006 Oct 18

[43] Bellew HW. An Inquiry into the Ethnography of Afghanistan. London: The Oriental University Institute, Woking; 1891

[44] Hamamy HA, Al-Allawi NA. Epidemiological profile of common haemoglobinopathies in Arab countries. Journal of Community Genetics. 2013;**4**(2):147-167. DOI: 10.1007/s12687-012-0127-8. Epub 2012 Dec 8

[45] Varawalla NY, Old JM, Sarkar R, Venkatesan R, Weatherall DJ. The spectrum of beta-thalassaemia mutations on the Indian subcontinent: The basis for prenatal diagnosis. British Journal of Haematology. 1991;**78**(2):242-247. DOI: 10.1111/j.1365-2141.1991.tb04423.x

[46] Varawalla NY, Fitches AC, Old JM. Analysis of beta-globin gene haplotypes in Asian Indians: Origin and spread of beta-thalassaemia on the Indian subcontinent. Human Genetics. 1992;**90**(4):443-449. DOI: 10.1007/BF00220475

[47] Haldane JBS. Disease and evolution. Ricerca Science Supplement. 1949;**19**:3-10

- [48] Roberts DJ, Williams TN. Haemoglobinopathies and resistance to malaria. *Redox Report*. 2003;**8**(5):304-310. DOI: 10.1179/135100003225002998
- [49] Cao A, Moi P. Regulation of the globin genes. *Pediatric Research*. 2002;**51**(4):415-421
- [50] Nordermeer D, de Laat W. Joining the loops: Beta-globin gene regulation. *IUBMB Life*. 2008;**60**(12):824-833
- [51] Ho PJ. The regulation of β globin gene expression and β thalassemia. *Pathology*. 1999;**31**:315-324
- [52] Ikonomi P, Noguchi CT, Miller W, Kassahun H, Hardison R, Schechter AN. Levels of GATA-1/GATA-2 transcription factors modulate expression of embryonic and fetal hemoglobins. *Gene*. 2001;**261**(2):277-287
- [53] Ropero P, Erquiaga S, Arrizabalaga B, Perez G, de la Iglesia S, Torrejon MJ, et al. Phenotype of mutations in the promoter region of the β -globin gene. *Journal of Clinical Pathology*. 2017;**70**(10):874-878
- [54] Giglioni B, Comi P, Ronchi A, Mantovani R, Ottolenghi S. The same nuclear proteins bind the proximal CACCC box of the human beta-globin promoter and a similar sequence in the enhancer. *Biochemical Biophysical Research Communications*. 1989;**164**(1):149-155
- [55] Siddoo-Atwal C. Genes that can cause cancer. In: *Gene Expression Profiling in Cancer*. London, UK: IntechOpen; 2019. DOI: 10.5772/intechopen.83549
- [56] Kukreti R, B-Rao C, Das SK, De M, Talukder G, Vaz F, et al. Study of the single nucleotide polymorphism (SNP) at the palindromic sequence of hypersensitive site (HS) 4 of the human β -globin locus control region (LCR) in Indian population. *American Journal of Hematology*. 2002;**69**:77-79
- [57] Chandra RG. *Indo-Greek Jewellery*. New Delhi: Abhinav Publications; 1979
- [58] Widemann F. Maues king of Taxila: An Indo-Greek kingdom with a Saka king. *East and West*. 2003; **53**(1/4): 95-125
- [59] Bernard P. In: Harmatta J, editor. Chapter 4: The Greek Kingdoms of Central Asia; in “History of Civilizations of Central Asia”. Paris, France: UNESCO Publishing; 1994. ISBN 978-92-3-102846-5
- [60] Siddoo-Atwal C. A role for heavy metal toxicity and air pollution in respiratory tract cancers. In: *Heavy Metal Toxicity in Public Health*. London, UK: IntechOpen; 2020. DOI: 10.5772/intechopen.90092
- [61] Most of the World Breathes Unsafe Air, Taking More Than 2 Years Off Global Life Expectancy. *Epic News*. 2022. Available from: <https://epic.uchicago.edu/news/most-of-the-world-breathes-unsafe-air-taking-more-than-2-years-off-global-life-expectancy/> [Accessed: September 12, 2024]
- [62] New Jersey Department of Health and Senior Services. Nitrogen Dioxide. Hazardous Substance Fact Sheet. New Jersey Department of Health and Senior Services; 2000. Available from: <https://nj.gov/health/eoh/rtkweb/documents/fs/1376.pdf> [Accessed: September 12, 2024]
- [63] California Air Resources Board. Carbon Monoxide and Health. California Air Resources Board; 2024. Available from: <https://ww2.arb.ca.gov/resources/carbon-monoxide-and-health> [Accessed: September 12, 2024]
- [64] American Lung Association. Ozone. American Lung Association; 2024.

Available from: <https://www.lung.org/clean-air/outdoors/what-makes-air-unhealthy/ozone> [Accessed: September 12, 2024]

[65] Agency for Toxic Substances and Disease Registry (US). Toxicological Profile for Sulfur Dioxide. Atlanta (GA): Agency for Toxic Substances and Disease Registry (US); 1998. Health Effects. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK596165/> [Accessed: September 12, 2024]

[66] Reuters. Almost 400,000 Deaths in Europe in 2021 Attributable to Filthy Air – EU Report. Reuters; 2023. Available from: <https://www.reuters.com/world/europe/almost-400000-deaths-europe-2021-attributable-filthy-air-eu-report-2023-11-24/> [Accessed: September 12, 2024]

[67] Kumar R, Singh K, Panigrahi I, Agarwal S. Genetic heterogeneity of beta globin mutations among Asian-Indians and importance in genetic counselling and diagnosis. *Mediterranean Journal of Hematology and Infectious Diseases*. 2013;5(1):e2013003. DOI: 10.4084/MJHID.2013.003. Epub 2013 Jan 2

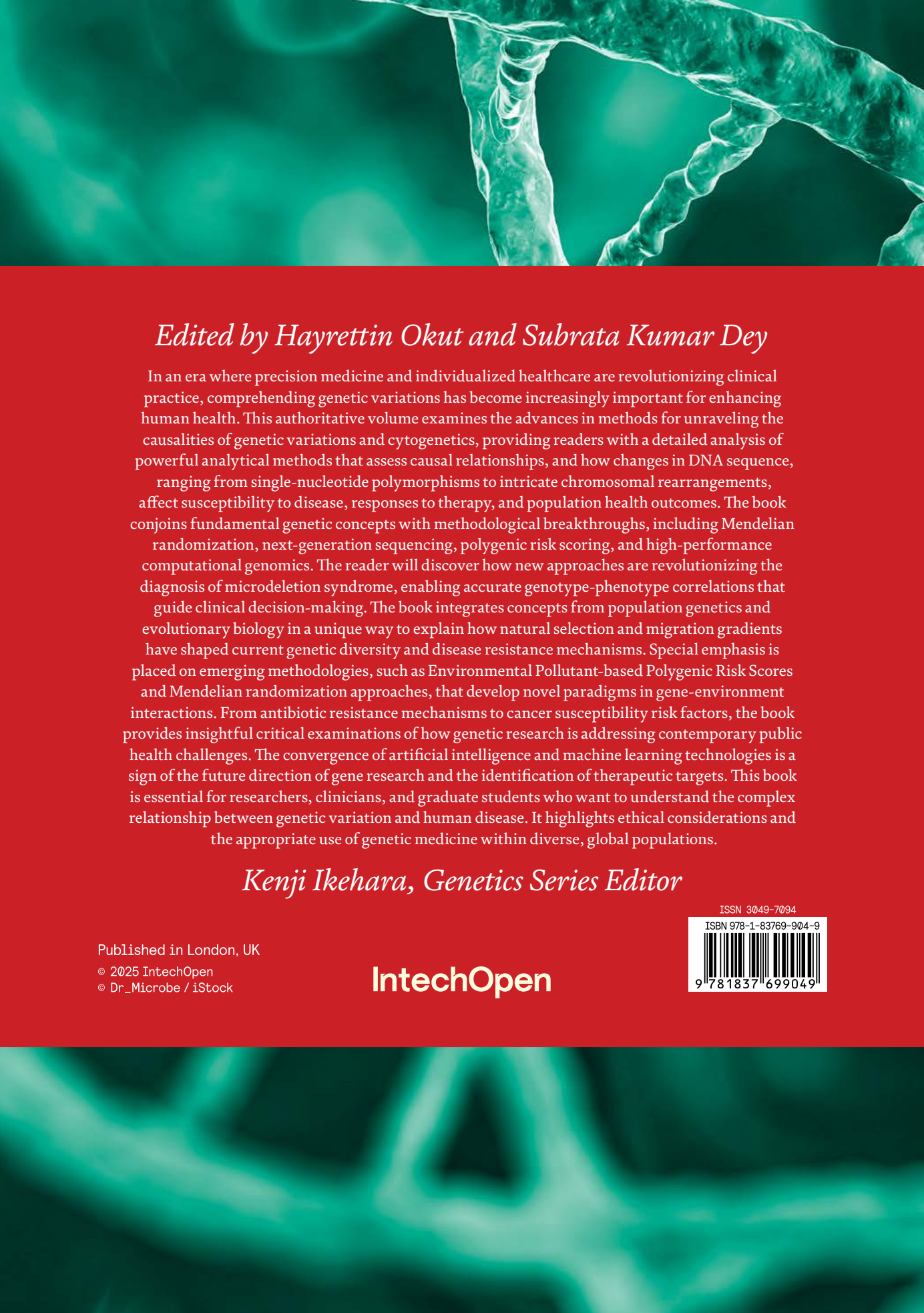
[68] Alkan C, Kavak P, Somel M, Gokcumen O, Ugurlu S, Saygi C, et al. Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe, Asia and Africa. *BMC Genomics*. 2014;15(1):963. DOI: 10.1186/1471-2164-15-963

[69] Akhavan-Niaki H, Derakhshandeh-Peykar P, Banihashemi A, Mostafazadeh A, Asghari B, Ahmadifard MR, et al. A comprehensive molecular characterization of beta thalassemia in a highly heterogeneous population. *Blood Cells, Molecules & Diseases*. 2011;47(1):29-32.

DOI: 10.1016/j.bcmed.2011.03.005. Epub 2011 Apr 13

[70] Ghirshman R. Iran. London, England: Penguin Publishing Group; 1978

[71] Strabo. *Strabo Geography* – Loeb Classical Library (Books 6-7; 15-16). London, England: Harvard University Press; 1995



Edited by Hayrettin Okut and Subrata Kumar Dey

In an era where precision medicine and individualized healthcare are revolutionizing clinical practice, comprehending genetic variations has become increasingly important for enhancing human health. This authoritative volume examines the advances in methods for unraveling the causalities of genetic variations and cytogenetics, providing readers with a detailed analysis of powerful analytical methods that assess causal relationships, and how changes in DNA sequence, ranging from single-nucleotide polymorphisms to intricate chromosomal rearrangements, affect susceptibility to disease, responses to therapy, and population health outcomes. The book conjoins fundamental genetic concepts with methodological breakthroughs, including Mendelian randomization, next-generation sequencing, polygenic risk scoring, and high-performance computational genomics. The reader will discover how new approaches are revolutionizing the diagnosis of microdeletion syndrome, enabling accurate genotype-phenotype correlations that guide clinical decision-making. The book integrates concepts from population genetics and evolutionary biology in a unique way to explain how natural selection and migration gradients have shaped current genetic diversity and disease resistance mechanisms. Special emphasis is placed on emerging methodologies, such as Environmental Pollutant-based Polygenic Risk Scores and Mendelian randomization approaches, that develop novel paradigms in gene-environment interactions. From antibiotic resistance mechanisms to cancer susceptibility risk factors, the book provides insightful critical examinations of how genetic research is addressing contemporary public health challenges. The convergence of artificial intelligence and machine learning technologies is a sign of the future direction of gene research and the identification of therapeutic targets. This book is essential for researchers, clinicians, and graduate students who want to understand the complex relationship between genetic variation and human disease. It highlights ethical considerations and the appropriate use of genetic medicine within diverse, global populations.

Kenji Ikehara, Genetics Series Editor

Published in London, UK

© 2025 IntechOpen
© Dr_Microbe / iStock

IntechOpen

ISSN 3049-7094

ISBN 978-1-83769-904-9



9 781837 699049