

Impact of Genetic Variation on Human Disease: From Genomics to Phenotype

Xiaohui Shi *

Northwest Women's and Children's Hospital, Xi'an Shaanxi, China
* Corresponding author Email: 2444385987@qq.com

Abstract: The impact of genetic variation on human disease is a multifaceted phenomenon involving genomics, epigenetics, and phenotype relationships. Advances in human genetic analysis technology have revolutionized our understanding of genetic disorders, enabling large-scale screens to explore associations between sequence variants and disease susceptibility. Environmental stimuli also contribute to epigenetic modifications that shape gene expression and phenotypic traits. Recent research on monozygotic twins highlights the role of non-genetic factors in phenotypic differences, underscoring the complex interplay between genetics and environment. The literature review highlights seminal studies that elucidate genetic and evolutionary mechanisms underlying human disease, providing insights for personalized medicine. Methods such as pharmacogenetics and immunomics provide avenues for precision therapies and immunological research. The Heterogeneity Weighted U test and the Gaussian random effects model facilitate the exploration of phenotype-genotype relationships and help to assess genetic heterogeneity in disease models. Simulation experiments deepen our understanding of the impact of genetic variation on disease manifestation. The interdisciplinary approaches are crucial to unravel the complex genetic landscape of human diseases and to develop tailored therapeutic interventions.

Keywords: Genetic Variation; Human Disease; Disease Susceptibility; Precision Medicine; Genetic Epidemiology.

1. Introduction

The relentless progress in human genetic information analysis technology, coupled with the finalization of the international human genome haplotype map, marks the inauguration of a novel epoch in human genetic inquiry. This breakthrough empowers large-scale genome-wide screenings

within populations, facilitating the elucidation of associations between sequence variants and disease susceptibility. Simultaneously, the advent of innovative genome exome sequencing techniques not only facilitates precise identification of both common and rare variants but also substantially curtails sequencing expenditures, thereby catalyzing extensive investigations into complex genetic disorders.

Table 1. The Impacts of Genetic Variation on Human Disease: Sample Demonstration

Category	Explanation	Illustrations
Single Nucleotide Differences (SNDs)	Variations in individual nucleotides, often impacting sole amino acids or control over gene expression	rs53576 (Oxytocin receptor gene)
Complex Gene-Related Diseases	Ailments caused by multiple gene variations interacting with environmental factors	Type 2 diabetes (T2D)
Uncommon Variances	Gene variations infrequent in the population yet potentially significant in specific diseases	Familial Mediterranean Fever (FMF)
Genomic Structural Changes	Involves structural modifications in the genome such as rearrangements of chromosomes, insertions, deletions, etc.	Chromosomal translocation in Chronic Myelogenous Leukemia (CML)
Variants in Coding Regions	Changes affecting the coding regions of genes, possibly leading to alterations in protein structure or function	Sickle cell anemia (HBB gene)
Variations in Regulatory Areas	Alterations affecting gene regulatory regions, influencing gene expression or binding of transcription factors	Non-coding variants in FTO gene associated with obesity
Epigenetic Alterations	Changes not involving modifications in DNA sequence but impacting gene expression or observable characteristics	DNA methylation changes in cancer
Genetic Modulators	Variations in one gene that influence the expression or effects of another gene	Modifier genes in cystic fibrosis

Environmental stimuli emerge as pivotal factors in the maturation and refinement of prenatal and early postnatal epigenetic mechanisms. The accumulation of misinformation

induced by these stimuli during epigenetic replication progressively engenders conspicuous epigenetic variations with advancing age. In recent years, the emergence of

environmental epigenomics has ushered in an era of exploration into the repercussions of environmental stimuli on epigenetic phenomena and gene expression. Investigations have revealed that under stress conditions, alterations occur in the DNA methylation patterns of plants, thereby modulating gene regulatory processes. This underscores the dynamic nature of the epigenome, which can be recalibrated based on the organism's experiences.

Collectively, these studies, alongside corroborating findings from other researchers, underscore the profound impact of life experiences on DNA, precipitating epigenetic alterations in humans. Notably, recent endeavors utilized whole-genome high-throughput analyses to scrutinize identical twins, focusing on a specific epigenetic modification—DNA methylation. Comparative analyses of DNA methylation levels among identical twins, non-affected twins, and those afflicted with autoimmune disorders unveiled substantial differences across numerous genes. The existence of such differences, despite nearly identical genetic compositions in identical twins, underscores the influence of non-genetic factors, such as environmental stimuli, in shaping phenotypic disparities among individuals.

These findings illuminate the multifaceted and intricate nature of human genetic exploration, furnishing novel perspectives and theoretical foundations for the comprehensive examination of the interplay between genetic inheritance and environmental influences. Moving forward, it is imperative to fortify research endeavors in genomics and epigenetics to garner a deeper understanding of the etiology of human genetic disorders and furnish more efficacious strategies and modalities for precision medicine and personalized therapeutic interventions. In the Table 1, the sample illustration of the impacts of genetic variation on human disease is presented.

For a comprehensive examination, the paper will delve into the influence of genetic variation on human disease, spanning from genomics to phenotype in the results section.

2. Literature Review

The literature review encompasses a diverse array of ten seminal studies spanning various domains, each offering profound insights into the complex genetic and molecular foundations of human ailments, the enduring impact of evolutionary dynamics on health, and the significant advancements in genomic research with implications for personalized medical interventions. Beginning with the groundbreaking work of Volpato and Webber [1], which meticulously examines the inherent variability within induced pluripotent stem cell (iPSC) models, proposing robust guidelines aimed at enhancing consistency in disease modeling experiments. Benton et al.'s comprehensive exploration [2] delves into the profound effects of evolutionary histories on human well-being, advocating for the integration of evolutionary frameworks into genetic medicine to craft customized healthcare modalities.

Furthermore, Wang et al.'s innovative investigation [3] meticulously unravels the contribution of infrequent protein-coding variations to human afflictions, shedding light on significant connections between genes and observable traits, underscoring the pivotal role of infrequent variations in shaping typical disease trajectories. Frazer et al.'s pioneering research [4] introduces EVE, an advanced deep generative model that predicts variant pathogenicity by leveraging evolutionary data, surpassing traditional methodologies and

offering nuanced insights for interpreting clinical variants.

Aganezov et al.'s groundbreaking discovery [5] of the Telomere-to-Telomere CHM13 (T2T-CHM13) genome heralds a new chapter in genomic research, elucidating its superiority over the GRCh38 reference genome and its potential to unlock novel aspects of human genetic variation. Zhang et al.'s pivotal findings [6] shed light on the intricate relationship between genetic variations in the IFITM3 gene and the severity of COVID-19, presenting innovative strategies for early intervention.

Moreover, Richardson et al.'s innovative approach [7] employs mendelian randomization to untangle the causal pathways linking early and later-life body size with disease susceptibility, providing nuanced insights into the complex interplay between physical dimensions and health outcomes. De Sá et al.'s groundbreaking inquiry [8] identifies genetic variations intricately linked with tuberculosis, HIV-1 infection, and TB/HIV-immune reconstitution inflammatory syndrome outcomes, illuminating the complex host-pathogen dynamics underlying these conditions.

Romeo et al.'s seminal perspective [9] explores the intricate genetic underpinnings of fatty liver disease, highlighting potential therapeutic targets derived from exhaustive human genetics studies. Lastly, Borgo et al.'s comprehensive review [10] delves into the multifaceted roles of the CK2 in various human ailments, emphasizing its potential as a pharmacological target across diverse medical domains.

Collectively, these seminal studies underscore the paramount importance of the genetic and evolutionary insights in elucidating the intricate mechanisms underlying human diseases, paving the way for personalized medicine and offering invaluable paradigms for disease diagnosis, treatment, and prevention in clinical settings.

3. The Proposed Methodology

3.1. The Genomics Disease Analysis

With the continuous progress in human genetic research, pharmacogenetics has become an indispensable and important part of medical practice. Its importance lies in guiding physicians to provide personalized drug treatment to patients and to assess the risk of adverse reactions that may occur during treatment. The core of pharmacogenetics is to study the changes in genes and RNA characteristics involved in drug treatment and the relationship between these changes and the diversity of drug metabolism, transport and target genes. By thoroughly analyzing the differences in drug effects among individuals, we can better understand the metabolism and mechanisms of action of drugs in the human body. When the results of pharmacogenetic research have an important impact on the safe use and effectiveness of a drug, this information is usually included in the relevant parts of the drug label, such as warnings, indications, dosage, contraindications, precautions, and drug-drug interactions. In this way, clinicians can better consider the impact of individual differences and genetic variation on treatment in practice. Pharmacogenetics not only provides an important tool for drug development and evaluation, but also helps us better understand the absorption, distribution, metabolism, excretion, and other processes of drugs in the human body, as well as the mechanism of action of drugs by detecting relevant markers. However, it should be noted that some pharmacogenetic markers involved in drug prescribing still lack approved detection tools for marketing, which is also one

of the important directions for future research and development. Immunomics is a discipline that studies the composition, structure, function and control mechanisms of the immune system under different conditions. It includes multiple branches such as antigenomics, immunogenomics, immunoproteomics, immunocytomics, antibodyomics and immunoinformatics. Antigenomics uses technologies such as peptide library, tetramer staining, and antibody library to explore the antigenic sites recognized by TCR or antibodies and study their functions and interaction methods. This helps in the design and development of immunotherapies or vaccines to treat diseases. Immunogenomics uses high-throughput sequencing technology to reveal the gene sequence of the human immune system and study the identification of immune-related genes, the correlation between diversity and disease susceptibility, and the regulatory mechanism of gene expression. Immunoproteomics studies the composition, structure and function of proteins in the immune system, analyzes changes in protein expression and modification, and discovers characteristic markers of immune-related diseases to provide support for prevention and treatment. Immunocytomics uses single-cell sequencing technology and flow cytometry to analyze the composition and functional characteristics of individual immune cells, study interactions between cells, and provide guidance for disease diagnosis and treatment. Antibodyomics accelerates the development of fully human antibody drugs by constructing large-scale antibody libraries and high-throughput screening technology to optimize antibodies for research, diagnosis or treatment. Immunoinformatics applies informatics methods in computer science and biology to immunological research, analyzing and interpreting data on the immune system, and revealing its structure, function and regulatory mechanisms. The development of immunomics provides new research directions and opportunities for interdisciplinary applications of immunology, genetics, and bioinformatics, helping to promote progress in human immune health and disease treatment. By deeply studying various aspects of the immune system in different states, we can better understand the nature of immune responses, discover early indicators of disease, and provide more effective strategies and means for personalized treatment and precision medicine.

3.2. The Phenotype Disease Relationships

Heterogeneity weighted U statistical model is selected a the modelling of the phenotype disease relationships. The Gaussian random effects model is used to explain random variation in data and is widely used, especially when exploring genetic correlations. This model takes into account random inter-individual differences that may exist in the data to more accurately assess the correlation between genes. In this model, the random effects are assumed to follow a Gaussian distribution, meaning that the random variation between individuals is normally distributed. This modeling approach allows us to account for inter-individual differences and is not limited to fixed effects analyses. As a result, we can better understand the differences between genes and their effects on phenotypic traits. Through the Gaussian random effects model, researchers can deeply analyze the correlation between genes and discover potential inheritance patterns and interactions between genes. This is critical for understanding the genetic mechanisms of disease, predicting individual risk, and developing personalized treatment plans. Therefore, the

Gaussian random effects model plays an important role in genetic research, providing us with the opportunity to deeply understand and unravel the complex relationship between genes and phenotypes defined as:

$$Y_i = \mu + Z_i^T \alpha + g_i \beta_i + \mathcal{R}_i \quad (1)$$

In the formula, Y_i : the phenotypic variable of the i -th individual;

Z_i : The covariate expected to be adjusted can be the main component for adjusting the population stratification;

g_i : single-site genotype random effect. g_i can take the values 0, 1 and 2 or 0 and 1;

β_i : parameters of random effects

$\mathcal{R}$: independent and identically distributed random errors

The Heterogeneity Weighted U Test is a statistical method that weights phenotypic similarity by genetic similarity. Its calculation process is a weighted sum of the phenotypic similarities of all individuals. Assuming that there is no correlation between phenotypic similarity and genetic similarity, the phenotypic similarity is zero, and therefore the expected value of the heterogeneity weighted U is also zero. Rejecting the null hypothesis and accepting the alternative hypothesis implies that as genetic similarity increases, phenotypic similarity also increases. This test gives greater weight to positive phenotypic similarities and less weight to negative phenotypic similarities, so that the heterogeneity-weighted U statistic has a positive value under the alternative hypothesis. The applicability of this test is reflected in its comprehensive consideration of the relationship between genetics and phenotype. By using genetic similarity as a weight, we are better able to assess changes in phenotypic similarity and thus understand the extent to which genetic factors influence phenotypic traits. If the alternative hypothesis is verified, it means that there is some degree of positive correlation between genetic similarity and phenotypic similarity among individuals in the study population. Applications of this statistical method include genetic research, disease genetic risk assessment, and population genetic structure analysis. By thoroughly exploring the relationship between genetics and phenotype, we can more accurately assess individual disease risks and provide more precise strategies for disease prevention and treatment.

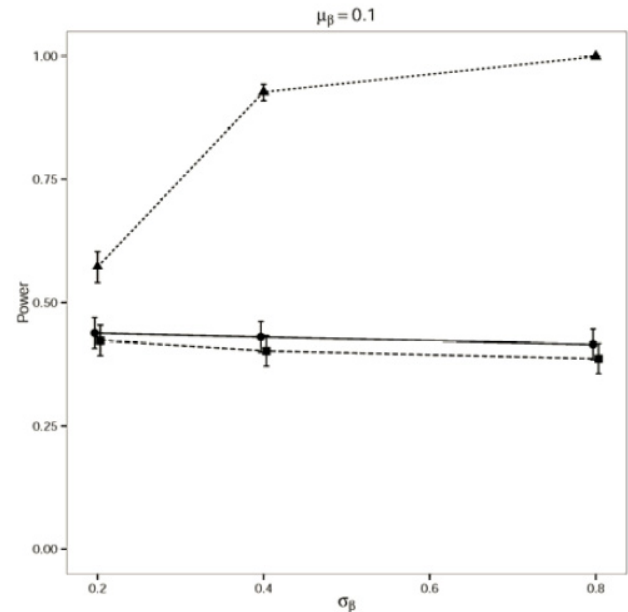


Figure 1. The Test Result for the Considered Scenario

In studies, genetic heterogeneity of a disease is often manifested when the effect of a genetic variant is above average. Therefore, the coefficient of variation is often used to evaluate the degree of genetic heterogeneity. In the simulation experiment, we set the mean and variance of the null model to zero, and gradually increased the coefficient of variation to simulate the disease model. The effectiveness of this approach lies in its ability to more accurately reflect the genetic nature of the disease. By gradually increasing the coefficient of variation, we can observe a trend toward a progressively increasing degree of genetic heterogeneity exhibited by the disease. Such simulation experiments can help us better understand the genetic mechanisms of disease and the extent to which genetic variation affects disease manifestations. Through simulation experiments, we can study the occurrence and development of genetic heterogeneity in more detail. This is important for identifying potential genetic factors, predicting the occurrence and development of disease, and formulating prevention and treatment strategies. The Figure 1 shows the experimental result.

4. Conclusion

The study sheds light on the intricate interplay between genetic variation and human disease, ranging from genomics to phenotype relationships. The relentless progress in genetic analysis technology has paved the way for comprehensive investigations into the genetic basis of disease. Epigenetic modifications and environmental cues further contribute to phenotypic diversity, emphasizing the need for interdisciplinary approaches to disease research. The literature review underscores the central role of genetic and evolutionary insights in disease understanding and provides paradigms for personalized medicine. Methods such as the pharmacogenetics and immunomics hold great promise for precision therapies and immunological research. The application of statistical models such as the Heterogeneity Weighted U test and the Gaussian random effects model helps to explore phenotype-genotype relationships and to assess genetic heterogeneity in disease. Simulation experiments provide valuable insights into the impact of genetic variation on disease manifestation. In the future, continued interdisciplinary research is essential to unravel the complexity of human diseases and develop targeted therapeutic strategies for personalized medicine.

References

- [1] Volpato, Viola, and Caleb Webber. "Addressing variability in iPSC-derived models of human disease: guidelines to promote reproducibility." *Disease models & mechanisms* 13, no. 1 (2020): dmm042317.
- [2] Benton, Mary Lauren, Abin Abraham, Abigail L. LaBella, Patrick Abbot, Antonis Rokas, and John A. Capra. "The influence of evolutionary history on human health and disease." *Nature Reviews Genetics* 22, no. 5 (2021): 269-283.
- [3] Wang, Quanli, Ryan S. Dhindsa, Keren Carss, Andrew R. Harper, Abhishek Nag, Ioanna Tachmazidou, Dimitrios Vitsios et al. "Rare variant contribution to human disease in 281,104 UK Biobank exomes." *Nature* 597, no. 7877 (2021): 527-532.
- [4] Frazer, Jonathan, Pascal Notin, Mafalda Dias, Aidan Gomez, Joseph K. Min, Kelly Brock, Yarin Gal, and Debora S. Marks. "Disease variant prediction with deep generative models of evolutionary data." *Nature* 599, no. 7883 (2021): 91-95.
- [5] Aganezov, Sergey, Stephanie M. Yan, Daniela C. Soto, Melanie Kirsche, Samantha Zarate, Pavel Avdeyev, Dylan J. Taylor et al. "A complete reference genome improves analysis of human genetic variation." *Science* 376, no. 6588 (2022): eabl3533.
- [6] Zhang, Yonghong, Ling Qin, Yan Zhao, Ping Zhang, Bin Xu, Kang Li, Lianchun Liang et al. "Interferon-induced transmembrane protein 3 genetic variant rs12252-C associated with disease severity in coronavirus disease 2019." *The Journal of infectious diseases* 222, no. 1 (2020): 34-37.
- [7] Richardson, Tom G., Eleanor Sanderson, Benjamin Elsworth, Kate Tilling, and George Davey Smith. "Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study." *Bmj* 369 (2020).
- [8] de Sá, Nathalia Beatriz Ramos, Nara Cristina Silva de Souza, Milena Neira-Goulart, Marcelo Ribeiro-Alves, Tatiana Pereira Da Silva, Jose Henrique Pilotto, Valeria Cavalcanti Rolla et al. "Inflammasome genetic variants are associated with tuberculosis, HIV-1 infection, and TB/HIV-immune reconstitution inflammatory syndrome outcomes." *Frontiers in Cellular and Infection Microbiology* 12 (2022): 962059.
- [9] Romeo, Stefano, Arun Sanyal, and Luca Valenti. "Leveraging human genetics to identify potential new treatments for fatty liver disease." *Cell metabolism* 31, no. 1 (2020): 35-45.
- [10] Borgo, Christian, Claudio D'Amore, Stefania Sarno, Mauro Salvi, and Maria Ruzzene. "Protein kinase CK2: a potential therapeutic target for diverse human diseases." *Signal Transduction and Targeted Therapy* 6, no. 1 (2021): 183.