Section: Miscellaneous

JAMP

Review Article

UNDERSTANDING COMPLEX TRAIT HERITABILITY: THE POLYGENIC APPROACH

Received : 26/02/2025
Received in revised form : 30/04/2025
Accepted : 19/05/2025

Keywords:
Polygenic risk scores, GWAS, Heritability, Pleiotropy, Systems genetics

Corresponding Author:
Dr. Uma Sankar Akula,
Email: uakula@tmsu.edu.vc

DOI: 10.47009/jamp.2025.7.3.135

Source of Support: Nil,
Conflict of Interest: None declared

Int J Acad Med Pharm
2025; 7 (3); 699-704

Uma Sankar Akula[1], Nagadharshan Devendra[1], Kelly Manahan., John Geisler[1]

[1]Trinity Medical Sciences University, Ratho Mill Campus, Saint Vincent and the Grenadines

ABSTRACT

Complex traits are governed by a multitude of genetic variants interacting across biological networks and influenced by environmental factors. This review explores the polygenic architecture of complex traits, emphasizing insights gained from genome-wide association studies (GWAS), polygenic risk scores (PRS), and systems genetics. It highlights the role of natural selection, regulatory variation, and population structure in shaping heritability patterns. The review also discusses emerging statistical models and omics integration techniques that enhance our ability to interpret polygenic signals. Finally, it addresses key challenges and future directions in understanding the molecular and evolutionary basis of complex trait heritability.

INTRODUCTION

Complex traits—such as height, intelligence, metabolic rate, and susceptibility to multifactorial conditions like type 2 diabetes, cardiovascular disease, and schizophrenia—are shaped by the interplay of numerous genetic loci and environmental stimuli. Unlike Mendelian disorders that result from single-gene mutations with large effect sizes and predictable inheritance patterns, complex traits follow a polygenic model of inheritance. In this model, hundreds or even thousands of genetic variants contribute to phenotypic variation, with each variant typically exerting only a small effect. These traits also tend to exhibit variable expressivity and incomplete penetrance due to modulation by environmental exposures, lifestyle factors, and epigenetic mechanisms. Over the past two decades, advancements in molecular genetics and statistical genomics have revolutionized our ability to dissect the genetic architecture of complex traits. The rise of genome-wide association studies (GWAS) has enabled researchers to scan the genome for associations between common single nucleotide polymorphisms (SNPs) and phenotypic variation across large population cohorts. These studies have uncovered thousands of loci associated with a broad spectrum of traits and diseases, supporting the notion of widespread polygenicity.[1] Furthermore, the development of polygenic risk scores (PRS) has introduced the potential to aggregate these small-effect variants into a cumulative genetic liability score, offering a predictive framework for individual risk stratification.

However, despite these technological breakthroughs, several important questions remain. Chief among these is the "missing heritability" problem—the observation that identified variants explain only a fraction of the heritability estimated from twin and family studies. In addition, interpreting the biological functions of associated SNPs remains challenging, especially when they reside in non-coding or intergenic regions. The predictive utility of PRS also varies across populations, highlighting the need for improved models that account for ancestry, gene–environment interactions, and functional regulatory mechanisms.[1,2] These challenges underscore the complexity of trait architecture and call for more integrative approaches that go beyond conventional association mapping.

The Concept of Polygenic Inheritance

The concept of polygenic inheritance has its roots in classical quantitative genetics. Ronald Fisher, in his seminal 1918 paper, introduced the idea that continuous traits such as height could be explained by the additive effects of numerous genes, each exerting a minute influence. This "infinitesimal model" reconciled Mendelian inheritance with continuous variation observed in the population by positing that the combined effect of many small-effect alleles would approximate a normal distribution.[1] This model laid the theoretical groundwork for much of modern complex trait genetics.

Empirical evidence supporting this model began to accumulate with the advent of GWAS, which provided genome-wide, hypothesis-free assessments of genotype–phenotype associations. The resulting discoveries confirmed that most complex traits are influenced by hundreds to thousands of variants, many of which fall outside traditional coding regions. For instance, traits such as height and body mass index have been linked to more than a thousand loci, each explaining only a small fraction of phenotypic variance. This diffuse pattern of effect sizes aligns well with Fisher's model and reaffirms that the genetic architecture of complex traits is highly polygenic.[2]

Moreover, the infinitesimal model has been extended to accommodate additional layers of complexity, including dominance, epistasis, gene–environment interactions, and pleiotropy. These extensions help explain not only trait variation but also evolutionary dynamics and the persistence of complex disease susceptibility alleles in the population. Together, they support the view that understanding complex traits requires integrating population genetics, molecular biology, and computational modeling.

The Omnigenic Model

While the polygenic model has been invaluable in advancing trait genetics, it has also raised important questions about the breadth of genetic involvement in complex traits. One such question is whether only genes directly involved in a trait's biological pathway are relevant, or whether a much larger set of genes may contribute indirectly. The omnigenic model, introduced by Boyle, Li, and Pritchard in 2017, addresses this by proposing that essentially all genes expressed in trait-relevant cells could influence complex traits through regulatory networks.[3]

According to the omnigenic hypothesis, core genes are those directly involved in biological processes governing the trait, while peripheral genes exert indirect effects by influencing the expression or function of these core genes through regulatory pathways. Because gene regulatory networks are dense and interconnected, perturbations in even distant loci can propagate effects across the network, thereby affecting core processes. This model provides a mechanistic explanation for why GWAS often implicate regions far from canonical genes, and why even genes with no obvious functional relationship to the trait of interest may still influence its expression.

The implications of the omnigenic model are profound. It suggests that heritability is not only widely dispersed but also deeply embedded within the architecture of cellular networks. This adds another layer of complexity to efforts aimed at pinpointing causal variants and developing precision medicine tools. Furthermore, it emphasizes the need for integrating GWAS with transcriptomic, epigenomic, and proteomic data to better map gene regulatory landscapes and their phenotypic consequences. In doing so, the omnigenic model reframes our understanding of genetic causality from a linear to a network-based perspective, highlighting the interconnectedness of the genome in shaping complex traits.

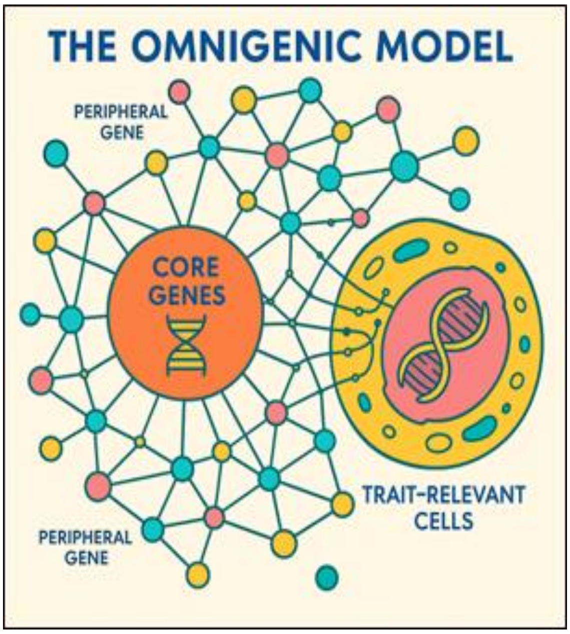

Figure 1: The Omnigenic Model

GWAS and the Discovery of Trait-Associated Loci

Over the past decade, genome-wide association studies (GWAS) have emerged as a cornerstone in uncovering the genetic underpinnings of complex traits. These studies scan the genomes of large populations to identify associations between single nucleotide polymorphisms (SNPs) and phenotypic variation, often without any prior hypothesis. To date, GWAS have identified thousands of loci linked to a wide variety of traits and diseases, including schizophrenia, type 2 diabetes, coronary artery disease, and educational attainment.[2] These findings have significantly expanded our catalog of genetic associations and reinforced the concept of polygenicity.

One striking observation from GWAS is the wide genomic dispersion of trait-associated loci. Many of these loci are found in non-coding regions, suggesting roles in gene regulation rather than direct protein alteration. Despite the large number of SNPs identified, each typically accounts for a minuscule portion of the phenotypic variance, often less than 0.1%. Collectively, even hundreds of these SNPs may explain only a fraction of the total heritability estimated through family or twin studies—a discrepancy known as the "missing heritability" problem.[4]

Several explanations have been proposed for this gap. First, GWAS primarily detect common variants and are underpowered to identify rare variants of potentially larger effect. Second, gene–gene (epistatic) and gene–environment interactions are not well captured in traditional GWAS frameworks. Third, structural variants, epigenetic modifications, and somatic mutations remain underexplored.

ISSN (O): 2687-5365; ISSN (P): 2753-6556

Consequently, while GWAS have been invaluable in mapping genetic architecture, they represent just one layer in the broader effort to understand complex trait heritability.

Polygenic Scores and Predictive Modeling

To address the highly polygenic nature of complex traits, researchers have developed polygenic scores—also referred to as polygenic risk scores (PRS)—as a means to quantify an individual's genetic predisposition. These scores are calculated by summing the effects of numerous SNPs, each weighted by its effect size derived from GWAS summary statistics.[5] PRS provide a continuous measure of genetic liability and are increasingly used in risk stratification, early screening, and personalized prevention strategies, especially in diseases like cardiovascular disorders, breast cancer, and psychiatric conditions. Despite their promise, the utility of PRS is constrained by several limitations. First, their predictive power is modest for most traits, with the exception of highly heritable and well-studied conditions. Second, PRS derived from GWAS conducted in European ancestry populations often perform poorly in non-European populations, due to differences in allele frequencies, linkage disequilibrium patterns, and population structure.[5,6] This raises concerns about the equity and generalizability of PRS-based clinical tools, especially in globally diverse populations.

Furthermore, PRS do not account for environmental exposures, lifestyle, or epigenetic modifications—factors that significantly modulate disease risk. There are also conceptual concerns about their interpretation: a high polygenic score reflects statistical association, not deterministic causality. To enhance their clinical relevance, ongoing research is focused on integrating PRS with additional data layers such as family history, environmental risk factors, and biomarkers. Multivariate and machine learning-based models are also being explored to better leverage the complex interdependencies among genetic predictors.

Role of Natural Selection in Polygenicity

While GWAS and PRS highlight the extensive spread of genetic contributions to complex traits, an important evolutionary question arises: why do so many small-effect variants persist in the population? One compelling explanation lies in the role of natural selection. O'Connor et al. argue that extreme polygenicity arises, in part, from the action of negative selection, which limits the frequency and effect size of deleterious alleles that influence fitness-related traits.[7] This results in a genetic architecture where many small-effect variants, rather than a few large-effect ones, contribute to phenotypic diversity. Negative selection tends to purge harmful alleles with large effects from the gene pool, especially if they influence traits that impact reproductive success or survival. However, variants with smaller effects can persist, accumulate, and collectively shape complex traits. This evolutionary constraint explains why GWAS often fail to detect high-effect loci for common diseases, despite extensive sample sizes. Instead, the architecture appears "flattened," with many loci contributing modestly to trait variance, making detection and prediction statistically challenging.

Moreover, traits that are tightly linked to evolutionary fitness, such as cognitive ability or reproductive timing, may exhibit higher degrees of polygenicity due to stronger purifying selection. This has implications not only for understanding trait biology but also for interpreting GWAS signals and refining models of heritability. It also underscores the importance of considering evolutionary dynamics when designing studies and interpreting the distribution of effect sizes across the genome.

Functional Interpretation of GWAS Hits

Despite the statistical success of GWAS in identifying thousands of trait-associated loci, the functional interpretation of these associations remains a major bottleneck in complex trait genetics. A notable proportion of significant SNPs from GWAS are found in non-coding regions of the genome, including intergenic areas, introns, and untranslated regions. These findings suggest that many GWAS signals likely exert their effects by modulating gene expression or chromatin state, rather than through direct changes to protein structure or function.[10]

To elucidate these regulatory mechanisms, researchers have increasingly relied on integrative functional genomics approaches. Expression quantitative trait loci (eQTL) mapping helps connect non-coding variants to downstream gene expression changes, while chromatin accessibility assays such as ATAC-seq and DNase-seq identify active regulatory elements in specific tissues or developmental contexts. These data sources, in combination with transcriptomic and proteomic profiles, allow for the prioritization of candidate genes and the construction of regulatory pathways linked to disease.[11]

Moreover, functional annotation tools and machine learning algorithms are being developed to predict the impact of non-coding variants based on sequence features and epigenomic signals. While these methods have improved our capacity to interpret GWAS findings, significant challenges remain. Tissue specificity, context-dependence, and cell-type heterogeneity all complicate the extrapolation of functional data to organism-level phenotypes. Bridging this gap will require deeper integration of multi-omics datasets and high-throughput experimental validation strategies.

Molecular Architecture and Regulatory Networks

The complexity of trait heritability extends beyond individual genes and variants to the broader molecular and regulatory architecture of the genome. Traits are not controlled by isolated genes but by dynamic networks of interacting transcripts, proteins, and regulatory elements. Gene expression is regulated by promoters, enhancers, transcription factors, non-coding RNAs, and chromatin-modifying enzymes—all of which can be influenced by genetic

ISSN (O): 2687-5365; ISSN (P): 2753-6556

variation. These interactions form intricate molecular networks that underlie phenotypic expression and contribute to the variability observed in complex traits.[12]

Lappalainen et al. have emphasized the importance of integrating multiple layers of molecular data to understand how genetic variation translates into phenotypic outcomes. For example, a single SNP might not directly cause a disease phenotype but may alter the expression of a transcription factor that controls an entire cascade of downstream genes. Understanding these intermediary steps is essential for decoding the full chain of causality between genotype and phenotype. Furthermore, different tissues may exhibit distinct regulatory landscapes, adding another dimension of complexity when linking genetic variation to disease processes.

Advances in high-throughput technologies, such as single-cell RNA sequencing and CRISPR-based perturbation assays, are now making it possible to systematically map these regulatory networks at unprecedented resolution. By integrating data from transcriptomics, epigenomics, proteomics, and spatial genomics, researchers can construct detailed models of gene regulation that reveal how multiple molecular mechanisms converge to influence complex traits. This systems-level understanding is not only critical for basic science but also for the development of more targeted and effective therapeutic strategies.

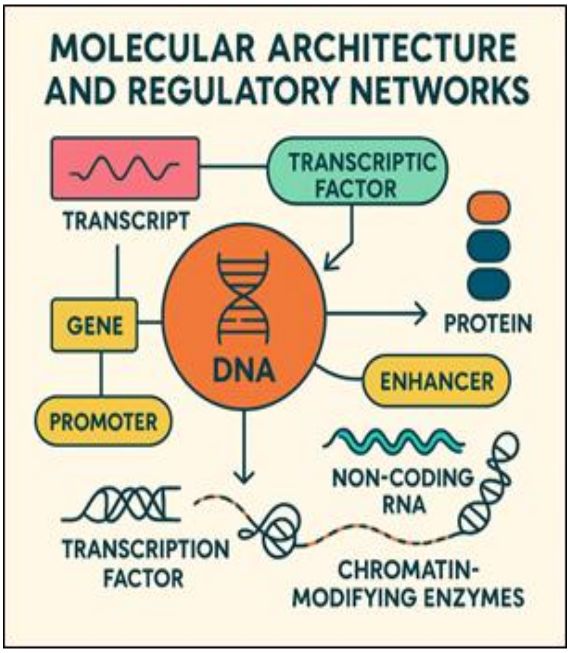

Figure 2: Molecular Architecture and Regulatory Networks

Statistical Models and Linear Mixed Approaches

As the field of complex trait genetics has evolved, so too have the statistical models used to interpret high-dimensional genomic data. Linear mixed models (LMMs) have become indispensable in GWAS for their ability to account for both population structure and polygenic background.[13] Unlike traditional linear regression models, LMMs incorporate a random effects component to capture the additive effects of genome-wide SNPs. This helps prevent false positives due to cryptic relatedness or ancestry stratification and yields more accurate estimates of SNP-based heritability.

LMMs have proven especially useful in large, ethnically diverse cohorts where uncorrected stratification could otherwise confound results. Moreover, their flexibility enables application to a wide variety of study designs, including case-control studies, longitudinal datasets, and repeated measures. Recent innovations include Bayesian mixed models, sparse LMMs, and models tailored for binary or time-to-event outcomes—each designed to improve scalability, interpretability, or computational efficiency.

Beyond GWAS, LMMs are now being extended to multi-omics contexts, integrating transcriptomic, proteomic, and epigenomic data to model how genetic variants influence downstream molecular phenotypes. This expansion allows for more precise trait mapping and improved causal inference. Similarly, LMMs are increasingly used in gene–environment interaction studies and for partitioning heritability across functional annotations or genomic regions, enabling a deeper understanding of the biological pathways involved in complex traits.

Systems Genetics and Integrative Omics

While single-variant analyses remain a mainstay in genetic studies, they are insufficient to fully capture the multifaceted nature of complex trait architecture. Systems genetics has emerged as a powerful framework that integrates various layers of omics data—including genomics, transcriptomics, epigenomics, metabolomics, and proteomics—within a network-based model.[14] The central tenet is that phenotypes arise not from isolated genes, but from the coordinated function of gene networks and molecular pathways influenced by both genetic and environmental variation.

Systems genetics enables the identification of key regulatory nodes, such as transcription factors or hub genes, that mediate the effects of genetic variants on downstream biological processes. These regulators are often the most biologically informative and the most promising targets for therapeutic intervention. For example, genetic variants may affect transcript levels in one tissue, leading to altered protein levels, cellular phenotypes, and eventually clinical manifestations. Mapping these cascades can uncover causal pathways and refine the interpretation of GWAS signals.

Multi-omics integration is further enhanced by computational tools such as co-expression network analysis, causal inference testing, and machine learning algorithms that model high-order interactions. The application of these methods has led to the discovery of trans-eQTLs, condition-specific regulatory effects, and multi-trait genetic correlations. Systems genetics not only deepens our understanding of complex trait biology but also holds promise for biomarker discovery, drug repurposing, and personalized medicine.

Population Structure and Bottleneck Effects

Understanding the influence of population history is essential in interpreting genetic association studies. Populations that have experienced demographic events such as bottlenecks, founder effects, or rapid population expansion can show altered patterns of allele frequency, linkage disequilibrium (LD), and genetic diversity. These factors affect both the power and accuracy of GWAS and polygenic risk score (PRS) models.[6]

In bottlenecked populations—such as the Finnish or Ashkenazi Jewish cohorts—alleles that are rare in the general population may reach higher frequencies, facilitating the discovery of trait-associated loci. However, the same demographic forces can also skew LD structure, complicating fine-mapping and replication efforts in other populations. Moreover, PRS derived from one population often show diminished predictive accuracy when applied to genetically distant groups, underscoring the need for more diverse and representative GWAS datasets.

To address these issues, researchers have begun incorporating demographic models and local ancestry inference into statistical pipelines. Additionally, efforts like the H3Africa project and the Global Biobank Meta-analysis Initiative are expanding the ancestral diversity of genomic databases. This is not only ethically important for equitable healthcare but also scientifically necessary to ensure that findings are generalizable and that genetic discoveries reflect the full spectrum of human diversity.

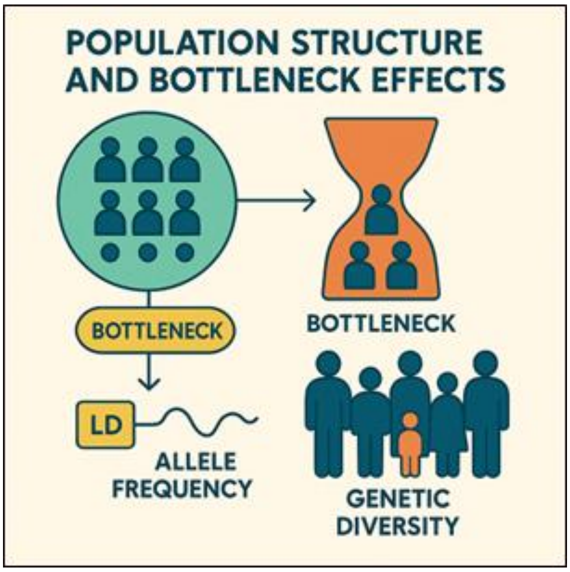

Figure 3: Population Structure and Bottleneck Effects

Future Directions and Open Challenges

Despite significant advances in uncovering the genetic basis of complex traits, numerous challenges remain that will shape the next phase of genomic research. One of the foremost challenges is the limited interpretability of polygenic risk scores. While PRS offer population-level risk stratification, their predictive power varies across traits and populations, and they often lack clinical utility on an individual level.[15] Enhancing PRS performance will require integrating environmental factors, rare variants, and multi-omic data into composite risk models.

Another major hurdle is fine-mapping causal variants among the vast sea of statistically associated SNPs. Improved resolution requires larger sample sizes, better functional annotations, and tissue- or cell-type–specific regulatory maps. This is especially crucial as most GWAS loci reside in non-coding regions, making biological inference difficult without additional functional data.

The integration of the omnigenic model further complicates interpretation by suggesting that peripheral genes—previously considered irrelevant—may significantly influence core biological processes through network effects.[16] This reframes heritability as a distributed property of gene regulatory networks rather than a function of a small set of trait-specific genes. Additionally, emerging evidence on gene–environment interactions, epigenetic memory, and microbiome-host dynamics point to a more layered and dynamic view of heritability.[17]

CONCLUSION

The polygenic framework has significantly deepened our understanding of complex trait heritability, revealing that numerous small-effect variants contribute to phenotypic variation across the genome. With advances in GWAS, polygenic risk scoring, and multi-omics integration, researchers are beginning to unravel the regulatory networks underlying trait expression. However, challenges such as missing heritability, population bias, and pleiotropy persist. Future research must incorporate functional genomics, evolutionary theory, and diverse populations to enhance predictive power and biological interpretation. A systems-level approach will be key to translating genetic discoveries into clinical and public health applications.

REFERENCES

1. Barton NH. Pleiotropic models of quantitative variation. Genetics. 1990 Mar;124(3):773–82. doi: 10.1093/genetics/124.3.773.
2. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017 Jul 6;101(1):5–22. doi: 10.1016/j.ajhg.2017.06.005. PMID: 28686856; PMCID: PMC5501872.
3. Ma Y, Zhou X. Genetic prediction of complex traits with polygenic scores: a statistical review. Trends Genet. 2021 Nov;37(11):995–1011. doi: 10.1016/j.tig.2021.06.004. PMID: 34243982; PMCID: PMC8511058.
4. O'Connor LJ, Schoech AP, Hormozdiari F, Gazal S, Patterson N, Price AL. Extreme polygenicity of complex traits is explained by negative selection. Am J Hum Genet. 2019 Sep 5;105(3):456–76. doi: 10.1016/j.ajhg.2019.07.003. PMID: 31402091; PMCID: PMC6732528.

5. Taylor CS, Lawson DJ. Heritability of complex traits in sub-populations experiencing bottlenecks and growth. J Hum Genet. 2024 Jul;69(7):329–35. doi: 10.1038/s10038-024-01249-2. PMID: 38589509; PMCID: PMC11199143.
6. Alamin M, Sultana MH, Lou X, Jin W, Xu H. Dissecting complex traits using omics data: a review on the linear mixed models and their application in GWAS. Plants. 2022;11(23):3277. doi: 10.3390/plants11233277.
7. Allayee H, Farber CR, Seldin MM, Williams EG, James DE, Lusis AJ. Systems genetics approaches for understanding complex traits with relevance for human disease. Elife. 2023 Nov 14;12:e91004. doi: 10.7554/eLife.91004. PMID: 37962168; PMCID: PMC10645424.
8. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. Cell. 2017 Jun 15;169(7):1177–86. doi: 10.1016/j.cell.2017.05.038. PMID: 28622505; PMCID: PMC5536862.
9. Lappalainen T, Li YI, Ramachandran S, Gusev A. Genetic and molecular architecture of complex traits. Cell. 2024 Feb 29;187(4):755–70.
10. Crouch DJM, Bodmer WF. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. Proc Natl Acad Sci U S A. 2020 Aug 4;117(32):18924–33. doi: 10.1073/pnas.2005634117.
11. Brandes N, Weissbrod O, Linial M. Open problems in human trait genetics. Genome Biol. 2022;23:131. doi: 10.1186/s13059-022-02697-9.
12. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, et al. Genetic effects on chromatin accessibility foreshadow gene expression changes in macrophage immune response. Nat Genet. 2018 Apr;50(4):515–23. doi: 10.1038/s41588-018-0046-7.
13. Barton NH, Etheridge AM, Veber A. The infinitesimal model. bioRxiv. 2016. doi: 10.1101/039768.
14. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, et al. Impact of regulatory variation from RNA to protein. Science. 2015 Feb 6;347(6222):664–7. doi: 10.1126/science.1260793.
15. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015 Sep;47(9):1236–41. doi: 10.1038/ng.3406. PMID: 26301638; PMCID: PMC4561502.
16. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet. 2003 Mar;33 Suppl:228–37. doi: 10.1038/ng1090.
17. Mathieson I. The omnigenic model and polygenic prediction of complex traits. Am J Hum Genet. 2021 Sep;108(9):1503–14. doi: 10.1016/j.ajhg.2021.07.003.