


## Estimation of Genetic Correlation Between Rheumatoid Arthritis and Multiple Sclerosis Using Summary Statistics from Genome-Wide Association Studies

 Ragıp Onur Öztornacı

Koç University Translational Research Center, Koç University, İstanbul, Türkiye

### ABSTRACT

**Objective:** Genome-wide association studies (GWAS) have revolutionized our understanding of the genetic basis of diseases by examining millions of genetic variants across the genome. Rheumatoid arthritis (RA) and multiple sclerosis (MS) are chronic autoimmune diseases characterized by immune system dysregulation and inflammation. Investigating the genetic correlation between RA and MS can provide insights into shared genetic factors, potential mechanisms, and pathways underlying these complex disorders. The objective of this study was to compare different statistical methods to estimate the genetic correlation between RA and MS using GWAS summary statistics.

**Materials and Methods:** To estimate single nucleotide polymorphism (SNP) heritability and genetic correlation, we utilized two popular methods: Linkage Disequilibrium Score Regression (LDSC) and Linkage Disequilibrium Adjusted Kinship (LDAK) models.

**Results:** Our analysis revealed a significant, moderate, positive correlation between RA and MS using both LDSC and LDAK ( $LDSC_{MS-RA}=0.448$ ,  $LDAK_{MS-RA}=0.387$ ,  $Spearman_{MS-RA}=0.0262$ ,  $p<0.001$ ). Additionally, there were notable differences in heritability estimates between the two methods and the traits. The LDAK model demonstrated higher heritability estimates for the RA-MS relationship ( $h^2_{MS-RA}=0.314$ ) compared to the LDSC ( $h^2_{RA-MS}=0.138$ ).

**Conclusion:** There is a significant positive genetic correlation between RA and MS, indicating a shared genetic component. Differential heritability estimates from LDAK and LDSC highlight the importance of the method. Genetic overlap informs common pathways and potential therapeutic targets. These findings contribute to the evidence of a moderately positive genetic correlation, emphasizing the need for further research and personalized approaches to managing autoimmune diseases.

**Keywords:** Genetic correlation, linkage disequilibrium score regression (LDSC), linkage disequilibrium adjusted kinship (LDAK), genome-wide association studies summary statistics, snp heritability.



#### Cite this article as:

Öztornacı RO. Estimation of Genetic Correlation Between Rheumatoid Arthritis and Multiple Sclerosis Using Summary Statistics from Genome-Wide Association Studies. J Clin Pract Res 2023; 45(6): 575–80.

#### Address for correspondence:

Ragıp Onur Öztornacı,  
Koç University Translational  
Research Center, Koç University,  
İstanbul, Türkiye  
**Phone:** +90 546 436 24 86  
**E-mail:**  
onur.oztornaci@gmail.com

**Submitted:** 17.07.2023

**Revised:** 21.08.2023

**Accepted:** 31.10.2023

**Available Online:** 04.12.2023

Erciyes University Faculty of  
Medicine Publications -  
Available online at [www.jcpres.com](http://www.jcpres.com)

### INTRODUCTION

Genome-wide association studies (GWAS) are a common and effective method for exploring the genetic architecture of diseases. GWAS rely on the differences in the frequencies of millions of



This work is licensed under  
a Creative Commons  
Attribution-NonCommercial  
4.0 International License.

genetic variants between patients and healthy controls. Summary statistics from GWAS, which provide estimates of the effect of each variant on risk, are now available for many different traits, often based on large sample sizes. An active area of current research is the statistical methodology for using summary GWAS data to extend our knowledge beyond the simple discovery of individual genetic variants. For example, summary statistics can be used to estimate heritability, and those from one trait can illuminate the genetic etiology of another trait by estimating the genetic correlation between them.<sup>1</sup> Autoimmune diseases are characterized by an aberrant immunological response in which the immune system mistakenly attacks the body's own cells and tissues. Normally, the immune system's function is to identify and eliminate harmful foreign entities such as bacteria and viruses. However, in autoimmune diseases, the immune system fails to distinguish between the body's own healthy cells and foreign invaders, leading to an immune reaction against self-antigens.<sup>2</sup> Multiple sclerosis (MS) and rheumatoid arthritis (RA) are chronic autoimmune conditions. However, they differ in terms of the organs affected and the specific mechanisms involved. In MS, the immune system targets and damages myelin, the protective layer surrounding nerve fibers in the central nervous system, which includes the brain and spinal cord. This leads to a disruption of nerve communication and a variety of neurological symptoms, such as fatigue, muscle weakness, balance and sensory disturbances, and coordination problems.<sup>2</sup> RA, conversely, primarily impacts the joints, resulting in inflammation, pain, and stiffness. The immune system's attack on the synovium, the tissue lining the joints, leads to joint damage and deformity over time. RA commonly affects the hands, wrists, feet, and knees but can also involve other organs and systems, causing systemic symptoms like fatigue, fever, and weight loss. While MS and RA have different target organs and manifestations, they share some similarities. Both diseases are driven by dysregulation of the immune system, with both genetic and environmental factors believed to play roles in their onset. Additionally, both conditions are chronic and require long-term management to control symptoms, slow disease progression, and preserve quality of life. Genetically, autoimmune diseases often exhibit associations with human leukocyte antigen (HLA) genes.<sup>3</sup> For instance, in RA, susceptibility to polymyalgia rheumatica is linked to HLA-DRB104 and DRB101 alleles. In MS, Human Leukocyte Antigen - DR Alpha chain (HLA-DRA), Interleukin 2 Receptor Alpha (IL2RA), High Mobility Group Box 1 (HMGB1), and Human Leukocyte Antigen - DQ Alpha 1 chain (HLA-DQA1) have been identified as significant factors.<sup>4,5</sup> These findings underscore the role of HLA genes in autoimmune disease susceptibility and provide insights into specific alleles associated with RA and MS. In our study, we utilized the GWAS catalog to obtain summary sta-

istics for investigating the genetic correlation between RA and MS. The GWAS catalog serves as a valuable resource that provides comprehensive information on genetic associations across various traits and diseases.<sup>6</sup>

## MATERIALS AND METHODS

### SNP Heritability

Single nucleotide polymorphisms (SNPs) can account for a portion of the phenotypic variance in various traits, and this proportion is referred to as SNP heritability. Estimating SNP heritability often involves variance component models fitted using restricted maximum likelihood (REML).<sup>7</sup> However, this approach may not be suitable for large biobanks such as the UK Biobank, which include a substantial number of individuals.<sup>8</sup> To address this issue, alternative methods and models have been developed to estimate SNP heritability. Commonly used approaches include Linkage Disequilibrium Score Regression (LDSC), Linkage Disequilibrium Adjusted Kinship (LDAK) models, and Genome-Wide Complex Trait Analysis (GCTA) models. LDSC and LDAK utilize summary statistics from genome-wide association studies (GWAS), which typically provide information such as *r*s numbers, alleles, sample sizes, and signed summary statistics (e.g., Z-scores) for LDSC or directionality (indicating a positive or negative effect) for LDAK. These approaches consider linkage disequilibrium (LD) patterns, which play a crucial role in understanding the genetic architecture of complex diseases. The LD patterns surrounding markers can affect the results, power, and complexity of GWAS studies, as they influence the association between diseases and alleles.<sup>9,10</sup> LD scores, derived from LD patterns, are utilized not only in calculating SNP heritability but also in computing Polygenic Risk Scores (PRS)<sup>11,12</sup> and estimating genetic correlations. It is worth noting that these methods rely on LD scores as a function of linkage disequilibrium to calculate SNP heritability. LD scores provide valuable insights into the genetic architecture of traits and diseases, allowing researchers to quantify the contribution of genetic factors and explore genetic correlations.

### Genetic Correlation

Genetic correlation is an estimation of the quantitative relationship between two different traits based on genetic similarity, which is affected by the same gene. This estimation can also reveal the contribution of the traits to biological causality.<sup>13–15</sup> The formula for calculating genetic correlation is as follows:

$$r_g = \frac{\partial_g}{\sqrt{h_1 h_2}} \quad (1)$$

where  $\partial_g$  is the genetic covariance, and  $h_1, h_2$  are the heritability estimates for the respective traits.<sup>16</sup>

### Genome-Wide Complex Trait Analysis (GCTA) Model

GCTA software can be used for several different purposes, such as data management, detecting genetic associations with SNPs, estimating LD structure, and simulating GWAS. GCTA uses the following equations to estimate total heritability:

$$A_{jk}^* = \begin{cases} 1 + \beta(A_{jj} - 1), & j = k \\ \beta(A_{jk}), & j \neq k \end{cases} \quad (2)$$

where  $\beta = 1 - (c + 1/N) / \text{var}(A_{jk})$ ,  $c + 1/N$  is the prediction error,  $\beta$  is a vector of fixed effects such as sex, age, and/or one or more eigenvectors from principal component analysis (PCA), and  $A$  is interpreted as the genetic relationship matrix (GRM) between individuals  $j$  and  $k$ . The GCTA model accounts for minor allele frequency (MAF) and extends the methodology to all chromosomes to estimate variance, allowing all SNPs to be used for variance estimation. This method requires individual-level PLINK files, meaning GCTA cannot be used with summary statistics.<sup>17</sup>

### Linkage Disequilibrium Score Regression (LDSC)

Linkage Disequilibrium Score Regression (LD score regression) is a reliable method based on GWAS summary statistics. LD score regression can describe separate polygenic effects and the contributions of confounding factors, such as population stratification, and can calculate correlations between different phenotypes.<sup>18,19</sup> LDSC requires summary statistics from large datasets, many of which are publicly available. LDhub is a useful database that includes many traits and provides an interface for using LDSC to calculate heritability and genetic correlation.<sup>19</sup>

LDSC uses the following equation (3) for genetic correlation between traits 1 and 2, with the expected value of  $z_{1j} z_{2j}$ -statistics for SNP  $j$  given by:

$$E[z_{1j} z_{2j} | l_j] = \frac{\sqrt{N_1 N_2} \partial_g}{M} l_j + \frac{\partial N_s}{\sqrt{N_1 N_2}} \quad (3)$$

where  $N_i$  is the sample size for study  $i$ ,  $\partial_g$  is the genetic covariance,  $l_j$  is the LD score of variant  $j$ ,  $N_s$  is the number of individuals included in both studies,  $\partial$  is the phenotypic correlation among  $N_s$  overlapping samples, and  $M$  is the number of variants.<sup>20</sup> The phenotypic correlation, which occurs due to sampling overlap, is a reason for false correlation inflation.

The LD Score for variant  $j$  can be defined as:

$$l_j = \sum_k r_{jk}^2 \quad (4)$$

where the sum is taken over all other SNPs  $k$ . In practice, there is very little LD in human samples outside of a small window, so LD Scores are typically estimated using a 1 centimorgan (cM) window. In the cases we consider, we will not need to calculate the intercept term  $(\frac{\partial N_s}{\sqrt{N_1 N_2}})$ , because the GWAS summary statistics we use do not come from overlapping samples.<sup>21</sup>

### Linkage Disequilibrium Adjusted Kinship (LDAK) Model

The LDAK model was developed to address the concept of LD, which posits that heritability is overestimated in regions of high LD and underestimated in regions of low LD. According to this concept, SNPs are weighted in relation to local LD, resulting in an LD-adjusted kinship matrix. Since LD patterns are strongly correlated with the MAF, it follows that heritability adjustments need to be made, with these adjustments depending on both LD and MAF. In other words, SNPs are weighted according to LD patterns and MAF.<sup>22</sup> LDAK uses the following formulation to calculate SNP heritability:

$$E[h_j^2] \propto [f_j(1 - f_j)]^{1+\alpha} * w_j * r_j \quad (5)$$

where  $E[h_j^2]$  represents the expected heritability contribution of SNP  $j$ ,  $f_j$  is its observed MAF, and  $\alpha$  is the parameter that determines the assumed relationship between heritability and MAF. General approaches do not account for MAF when calculating heritability, which is achieved by setting  $\alpha = -1$ ; however, LDAK considers alternatives. SNP weights ( $w_j$ ) are computed based on LD levels, and  $r_j \in [0,1]$  is a score reflecting the certainty of genotype information. In the LDAK model, this score inversely relates the quality of an SNP to its contributions.<sup>23</sup> To highlight the main differences between LDAK and LDSC, we can consider the following factor:

$$q_j = [f_j(1 - f_j)]^{0.75} * w_j \quad (6)$$

In LDAK,  $E[h_j^2] \propto q_j$ , meaning that the user can specify  $q_j$  arbitrarily, where  $f_j$  represents the MAF of SNP  $j$  and  $w_j$  is a weight based on local levels of LD. When estimating heritability with LDSC,  $q_j = 1$ , meaning that all SNPs contribute equally.<sup>24</sup>

Two popular methods for estimating SNP heritability and genetic correlation using GWAS summary statistics are the LDSC and LDAK models. We applied these models to estimate genetic correlation and SNP heritability.

### Datasets

The National Human Genome Research Institute (NHGRI) - European Bioinformatics Institute (EBI) GWAS Catalog, a database of genome-wide association studies, has been providing data from published studies since 2008. In 2015, the catalog underwent a redesign and transitioned to the European Molecular Biology Laboratory (EMBL) - European Bioinformatics Institute (EBI). The updated infrastructure includes a new user interface accessible at [www.ebi.ac.uk/gwas/](http://www.ebi.ac.uk/gwas/), improved search capabilities supported by ontology, and an enhanced curation interface.<sup>5</sup> Okada et al.<sup>25</sup> conducted a meta-analysis in 2014 to detect susceptibility SNPs for RA, based on data from >100,000 subjects of European and Asian ancestries, including 29,880 RA cases and 73,758 controls. Additionally, the European RA GWAS meta-analysis included 14,361 RA cases and 43,923 controls. We used a large GWAS summary statistics

**Table 1.** Results of genetic correlation and heritability

Description			SumHer (LDAK)			LDSC		Spe Cor	
Traits	Number of SNPs	Sample size (N)	Heritability (SD)	Correlation (SD)	CoHer (SD)	Heritability (SD)	Correlation (SD)	Z value	Spe Coe
RA	8,747,962	58,000	0.314 (0.026)	0.387 (0.116)	0.081 (0.024)	0.138 (0.021)	0.448 (0.169)	2.646	0.0262
MS	472,086	26,621							p<0.001
RA	8,747,962	58,000	0.015 (0.055)	0.002 (0.064)	0.061 (0.007)	0.014 (0.027)	0.006 (0.003)	0.083	0.0002
T2D	8,403,414	115,000							p=0.466
MS	472,086	26,621	0.022 (0.060)	0.010 (0.054)	0.018 (0.062)	0.0127 (0.148)	0.018 (0.339)	0.462	0.0006
T2D	8,403,414	115,000							p=0.678

CoHer: Coheritability; SD: Standard deviation; Spe Cor: Spearman correlation; Spe Coe: Spearman coefficient; SNPs: Single nucleotide polymorphisms; LDAK: Linkage disequilibrium adjusted kinship; LDSC: Linkage disequilibrium score regression.

dataset from the GWAS catalog for MS, comprising 9,772 cases and 17,376 controls, with a total of 472,086 SNPs.<sup>26</sup> The Type II Diabetes (T2D) dataset included data on obesity (29,925 cases) and type 2 diabetes (4,040 cases) among 120,286 individuals of British ancestry, obtained from the UK Biobank.<sup>27</sup>

## RESULTS

We converted GWAS summary statistics, which include Z Score, A1, A2, N, and P-values, into a text file for use with LDSC. Our analyses were conducted in Linux environments for both methods, as the LDhub online methods did not allow the configuration we required, specifically, the omission of an intercept term. We had to use the no-intercept term because there is no sample overlap between MS and RA. LDSC uses the HapMap3 SNP allele list when calculating LD scores<sup>27</sup> and can exclude major histocompatibility complex (MHC) regions to provide an accurate genetic correlation estimate. Additionally, we used summary statistics suitable for the LDAK model. When using the LDAK model, MHC was removed manually for all traits to avoid multicollinearity when calculating tagging files. We also applied the traditional Spearman Correlation method to estimate the correlation between Z-scores using data merged based on minor alleles.<sup>28</sup> SNPs with identical reference alleles for each trait were merged. The MHC region plays an important role in autoimmune diseases, being the strongest genetic risk factor for both RA and MS. To prevent biased heritability estimation, SNPs within the MHC, a region of extremely high LD, were excluded in LDSC as well as in each analysis. The exclusion of this region may have contributed to the low mean chi-squared values of the test statistics. LDSC is sensitive to the following criteria when calculating genetic correlation:

- Heritability (H2) Z score is at least >1.5 (ideally >4).
- The mean Chi-square of the test statistics >1.02.
- The intercept estimated from the SNP heritability analysis is between 0.9 and 1.1.

When utilizing LDSC in a Linux environment, we were unable to calculate a genetic correlation between MS and RA due to differences in sample sizes. Consequently, the LD-Hub online method also failed to yield any results. However, we successfully calculated heritability and genetic correlation for all other traits. We also used LDAK to estimate the correlation between MS and RA. LDAK incorporates several analysis methods, one of which is SumHer, used for calculating genetic correlations based on GWAS summary statistics.<sup>23,24</sup> Both approaches revealed a moderately positive correlation between MS and RA (LDSC<sub>MS-RA</sub>=0.448, LDAK<sub>MS-RA</sub>=0.387, Spearman<sub>MS-RA</sub>=0.0262, p<0.001). The heritability results demonstrated considerable differences between the two models and the traits analyzed. The highest heritability was observed for the RA-MS comparison using the LDAK model ( $h^2_{MS-RA}$ =0.314), followed by the RA-MS comparison using the LDSC model ( $h^2_{RA-MS}$ =0.138). When comparing the two methods, LDAK generally produced higher heritability results than LDSC. Our subsequent analyses, employing both LDSC and LDAK methods, found no statistically significant correlation between RA and Type 2 Diabetes (T2D) (LDSC<sub>RA-T2D</sub>=0.006, LDAK<sub>RA-T2D</sub>=0.002, Spearman<sub>RA-T2D</sub>=0.0002, p=0.466) or between MS and Type 2 Diabetes (LDSC<sub>MS-T2D</sub>=0.018, LDAK<sub>MS-T2D</sub>=0.010, Spearman<sub>MS-T2D</sub>=0.0006, p=0.678) (Table 1).

## DISCUSSION

In this study, our objective was to explore the genetic correlation between two diseases, MS and RA. We observed that the LDAK method tends to estimate higher heritability compared to the LDSC, which aligns with existing literature (Table 1).<sup>23</sup> The sample size played a significant role in calculating the genetic correlation, with LDSC being more affected than the LDAK model in our study. The LDAK model, incorporating linkage disequilibrium (LD)-weighted SNP contribu-

tions, yielded higher estimates than the LDSC model, which might be advantageous for heritability estimation. However, it is crucial to consider potential biases when calculating the genetic correlations.<sup>28</sup> Our findings demonstrate that LDAK heritability estimates were higher than those of LDSC, while the genetic correlation results did not exhibit substantial differences between the two models. It is important to note that estimating genetic correlation using GWAS summary statistics may not provide complete accuracy due to various factors, such as different populations, distinct LD patterns, and genotype-environment interactions.<sup>29</sup> Our study revealed a significant positive correlation using both the LDAK and LDSC methods. Although the MS study may have had limitations regarding the number of SNPs included, which could affect LDSC, both LDAK and LDSC methods yielded consistent results, indicating a moderately positive genetic correlation between RA and MS. Additionally, the classical approach, such as the Spearman correlation, was deemed unsuitable for calculating genetic correlation.<sup>19,29</sup> Incorporating Type 2 Diabetes into our study has broadened the scope of our investigation by introducing an unrelated condition with a distinct etiology from autoimmunity. Upon careful analysis of the data, it was not consistently observed that Type 2 Diabetes exhibited a significant correlation with either Multiple Sclerosis (MS) or Rheumatoid Arthritis (RA) using both the LDAK and LDSC methods. Consequently, it is plausible to consider that the observed genetic correlation between RA and MS may stem from their shared autoimmune basis.

## CONCLUSION

In this study, we found a strong positive genetic correlation between RA and MS, which are both autoimmune diseases. It is important to note that MHC regions were excluded from all traits and methods. Investigating the similarities and differences between RA and MS may help identify similarities or differences in treatment approaches. Some medications used to treat RA, such as certain disease-modifying antirheumatic drugs (DMARDs),<sup>30</sup> may also offer potential benefits or risks when considered in the context of MS, and vice versa. By studying both diseases together, researchers can gain insights into potential cross-utilization or shared therapeutic strategies. Understanding the relationship between RA and MS can help improve diagnostic criteria, enable earlier identification of comorbidities, and guide appropriate treatment decisions for individuals with overlapping symptoms. In cases where there is suspicion of correlation between genetic diseases, the use of Spearman's correlation coefficient can lead to misleading results. Therefore, when examining correlations within genetic data, it is advisable to use statistical methods such as LDAK and LDSC that take into account genetic patterns.

**Peer-review:** Externally peer-reviewed.

**Informed Consent:** Written informed consent was obtained from patients who participated in this study.

**Conflict of Interest:** The author have no conflict of interest to declare.

**Financial Disclosure:** The author declared that this study has received no financial support.

## REFERENCES

1. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nature Reviews Methods Primers* 2021; 59: 1–21. [\[CrossRef\]](#)
2. Pisetsky DS. Pathogenesis of autoimmune disease. *Nat Rev Nephrol* 2023; 19(8): 509–24. [\[CrossRef\]](#)
3. Barkhane Z, Elmadi J, Satish Kumar L, Pugalenthil LS, Ahmad M, Reddy S. Multiple Sclerosis and Autoimmunity: A Veiled Relationship. *Cureus* 2022; 14(4): e24294. [\[CrossRef\]](#)
4. Asouri M, Alinejad Rokni H, Sahraian MA, Fattahi S, Motamed N, Doosti R, et al. Analysis of single nucleotide polymorphisms in HLA-DRA, IL2RA, and HMGB1 genes in multiple sclerosis. *Rep Biochem Mol Biol* 2020; 9(2): 198–208. [\[CrossRef\]](#)
5. Thibaut D, Sweeney C, South S, Hussein M. Graves' disease and major histocompatibility complex class II: A meta-analysis of Hla-Dq And Hla-Drb1. *Advances in Clin Med Res Healthcare Delivery* 2023; 3(1): 2. [\[CrossRef\]](#)
6. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource. *Nucleic Acids Res* 2023; 51(D1): D977–85. [\[CrossRef\]](#)
7. Berrandou TE, Balding D, Speed D. LDAK-GBAT: Fast and powerful gene-based association testing using summary statistics. *Am J Hum Genet* 2023; 110(1): 23–9. [\[CrossRef\]](#)
8. Hou K, Burch KS, Majumdar A, Shi H, Mancuso N, Wu Y, et al. Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture. *Nat Genet* 2019; 51(8): 1244–51. [\[CrossRef\]](#)
9. Zhao B, Zou F, Zhu H. Cross-trait prediction accuracy of summary statistics in genome-wide association studies. *Biometrics* 2023; 79(2): 841–53. [\[CrossRef\]](#)
10. Ragsdale AP. Local fitness and epistatic effects lead to distinct patterns of linkage disequilibrium in protein-coding genes. *Genetics* 2022; 221(4): iyac097. [\[CrossRef\]](#)
11. Gazal S, Finucane HK, Furlotte NA, Loh PR, Palamara PF, Liu X, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet* 2017; 49(10): 1421–7. Erratum in: *Nat Genet* 2019; 51(8): 1295. [\[CrossRef\]](#)

12. Pham D, Truong B, Tran K, Ni G, Nguyen D, Tran TTH, et al. Assessing polygenic risk score models for applications in populations with under-represented genomics data: an example of Vietnam. *Briefings in Bioinformatics* 2022; 23(6): bbac459. [\[CrossRef\]](#)
13. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. An integrated framework for local genetic correlation analysis. *Nat Genet* 2022; 54(3): 274–82. [\[CrossRef\]](#)
14. Ling JJ, Ziao Y, Hu J, Wang F, Quyang F, Wang JH, et al. Genotype by environment interaction analysis of growth of *Picea koraiensis* families at different sites using BLUP-GGE. *New Forests* 2021; 52(1): 113–27. [\[CrossRef\]](#)
15. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: from theory to practice. *Nat Rev Genet* 2019; 20(10): 567–81. [\[CrossRef\]](#)
16. Wang M, Cao X, Zhang S, Sha Q. A clustering linear combination method for multiple phenotype association studies based on GWAS summary statistics. *Sci Rep* 2023; 13(1): 3389. [\[CrossRef\]](#)
17. Hunter MD, McKee KL, Turkheimer E. Simulated nonlinear genetic and environmental dynamics of complex traits. *Dev Psychopathol* 2023; 35(2): 662–77. [\[CrossRef\]](#)
18. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J; Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; 47(3): 291–5. [\[CrossRef\]](#)
19. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al; Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium; et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2017; 33(2): 272–9. [\[CrossRef\]](#)
20. Zhou W, Kanai M, Wu KH, Rasheed H, Tsuo K, Hirbo JB, et al. Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genom* 2022; 2(10): 100192. [\[CrossRef\]](#)
21. Elgart M, Goodman MO, Isasi C, Chen H, Morrison AC, de Vries PS, et al. Correlations between complex human phenotypes vary by genetic background, gender, and environment. *Cell Rep Med* 2022; 3(12): 100844. [\[CrossRef\]](#)
22. Speed D, Kaphle A, Balding DJ. SNP-based heritability and selection analyses: Improved models and new results. *Bioessays* 2022; 44(5): e2100170. [\[CrossRef\]](#)
23. Srivastava AK, Williams SM, Zhang G. Heritability estimation approaches utilizing genome-wide data. *Curr Protoc* 2023; 3(4): e734. [\[CrossRef\]](#)
24. Speed D, Balding DJ. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat Genet* 2019; 51(2): 277–84. [\[CrossRef\]](#)
25. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014; 506(7488): 376–81. [\[CrossRef\]](#)
26. International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2; Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; 476(7359): 214–9. [\[CrossRef\]](#)
27. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet* 2018; 50(7): 906–8. [\[CrossRef\]](#)
28. Lee SH, Yang J, Chen GB, Ripke S, Stahl EA, Hultman CM, et al. Estimation of SNP heritability from dense genotype data. *Am J Hum Genet* 2013; 93(6): 1151–5. [\[CrossRef\]](#)
29. Zhang Z, Ma P, Li Q, Xiao Q, Sun H, Olasege BS, et al. Exploring the genetic correlation between growth and immunity based on summary statistics of genome-wide association studies. *Front Genet* 2018; 9: 393. [\[CrossRef\]](#)
30. van Sleen Y, van der Geest KSM, Huckriede ALW, van Baarle D, Brouwer E. Effect of DMARDs on the immunogenicity of vaccines. *Nat Rev Rheumatol* 2023; 19(9): 560–75. [\[CrossRef\]](#)