PRS-Net: Interpretable polygenic risk scores via geometric learning

Han Li[1], Jianyang Zeng[2,*], Michael P. Snyder[3,*], and Sai Zhang[4,5,6,*]

[1] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China
[2] School of Engineering, Westlake University, Hangzhou, Zhejiang, China
[3] Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA
[4] Department of Epidemiology, University of Florida, Gainesville, FL, USA
[5] J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA
[6] The Genetics Institute, University of Florida, Gainesville, FL, USA
[*] Correspondence: zengjy@westlake.edu.cn; mpsnyder@stanford.edu; sai.zhang@ufl.edu

Abstract. Polygenic risk score (PRS) serves as a valuable tool for predicting the genetic risk of complex human diseases for individuals, playing a pivotal role in advancing precision medicine. Traditional PRS methods, predominantly following a linear structure, often fall short in capturing the intricate relationships between genotype and phenotype. We present PRS-Net, an interpretable deep learning-based framework designed to effectively model the nonlinearity of biological systems for enhanced disease prediction and biological discovery. PRS-Net begins by deconvoluting the genome-wide PRS at the single-gene resolution, and then it encapsulates gene-gene interactions for genetic risk prediction leveraging a graph neural network, thereby enabling the characterization of biological nonlinearity underlying complex diseases. An attentive readout module is specifically introduced into the framework to facilitate model interpretation and biological discovery. Through extensive tests across multiple complex diseases, PRS-Net consistently outperforms baseline PRS methods, showcasing its superior performance on disease prediction. Moreover, the interpretability of PRS-Net has been demonstrated by the identification of genes and gene-gene interactions that significantly influence the risk of Alzheimer's disease and multiple sclerosis. In summary, PRS-Net provides a potent tool for parallel genetic risk prediction and biological discovery for complex diseases.

1 Introduction

Complex human diseases display polygenicity in their genetic architectures, characterized by a multitude of common genetic variants with minor individual effects accumulatively influencing the disease risk[1–4]. The polygenic risk scores (PRSs) are developed to quantitatively characterize the genetic susceptibility of individuals to specific traits or complex diseases based on the common genetic variants[5–7]. This methodology empowers the early deployment of targeted therapeutic interventions and facilitates the practice of personalized medicine[8–10].

PRS is typically calculated using the summary statistics derived from genome-wide association studies (GWAS)[11–17], a widely-used statistical method for identifying disease-associated genetic variants[18–20]. While GWAS can identify disease risk genetic variants, such as single nucleotide polymorphisms (SNPs), that exhibit significant differences in frequencies between cases and controls, these variants tend to have modest individual effects on the phenotype, resulting in limited prediction capability. In an effort to enhance predictive modeling, various statistical methods have been applied to aggregate the effects of individual SNPs. The widely adopted method for calculating PRS, exemplified by PLINK[21] and PRSice[12], is known as clumping and thresholding (C+T)[11], which involves summing allele counts weighted by effect sizes estimated from GWAS. More recent approaches like LDpred2[16] utilize Bayesian modeling to infer the posterior mean effect size of each marker by incorporating prior information on effect sizes and linkage disequilibrium (LD) data from an external reference panel. Similarly, lassosum2[17] estimates PRS using summary statistics and a reference panel within a penalized regression framework. With the notable increase in dataset sample sizes for GWAS, these methods have achieved enhanced predictive power[22]. Nonetheless, these techniques primarily rely on univariate effect sizes derived from linear GWAS models, thus often overlook potential non-linear associations between genetic factors and phenotypes, which can undermine their predictive performance.

Efforts have also been made to construct models capable of capturing non-linear interactions in PRS calculation. These include tree-based methods like random forests[23,24], gradient boosting[25,26], and AdaBoost[27,28], as well as deep learning-based techniques such as multiple-layer perceptrons (MLP)[29] and convolutional neural networks[30]. However, these methods only take a limited number of variants as their input, and lack the integration of versatile prior biological knowledge. Indeed, these approaches have demonstrated either comparable or, in many cases, less effective performance in predicting phenotypes when compared to linear models[31,32].

In this study, we propose PRS-Net, a geometric deep learning-based approach designed to effectively model the intricate non-linear relationships among genetic factors such as genes in predicting the disease risk, thus delivering more accurate and robust PRSs. Based on the summary statistics of GWAS, PRS-Net first maps PRS onto a gene-gene interaction (GGI) network through the derivation of gene-level PRSs using the C+T method. Subsequently, a graph neural network is employed to iteratively update the embedding of the genes via performing message passing on the GGI network, thus capturing the complex GGIs from the network. An attentive readout module is then introduced to provide interpretable PRS predictions. PRS-Net also integrates a mixture-of-expert module[33] designed to enhance the accuracy of PRS predictions when dealing with multi-ancestry datasets. Our comprehensive evaluation encompasses six complex diseases extracted from the UK Biobank database[34], including Alzheimer's disease, atrial fibrillation, rheumatoid arthritis, multiple sclerosis, ulcerative colitis, and asthma. The results consistently demonstrated the superiority of PRS-Net over baseline methods, including PLINK[21], PRSice2[14], LDpred-2[16], and lassosum2[17] in PRS prediction. Notably, through case studies focused on Alzheimer's disease and multiple sclerosis, we illustrated that PRS-Net provided biologically meaningful interpretability by identifying specific genes and GGIs that significantly influence

disease risk. In summary, PRS-Net stands as a potent and innovative tool for precise PRS prediction, addressing the limitations of current linear models and offering a more comprehensive approach to unraveling the genetic underpinnings of complex traits and diseases.

2 Method

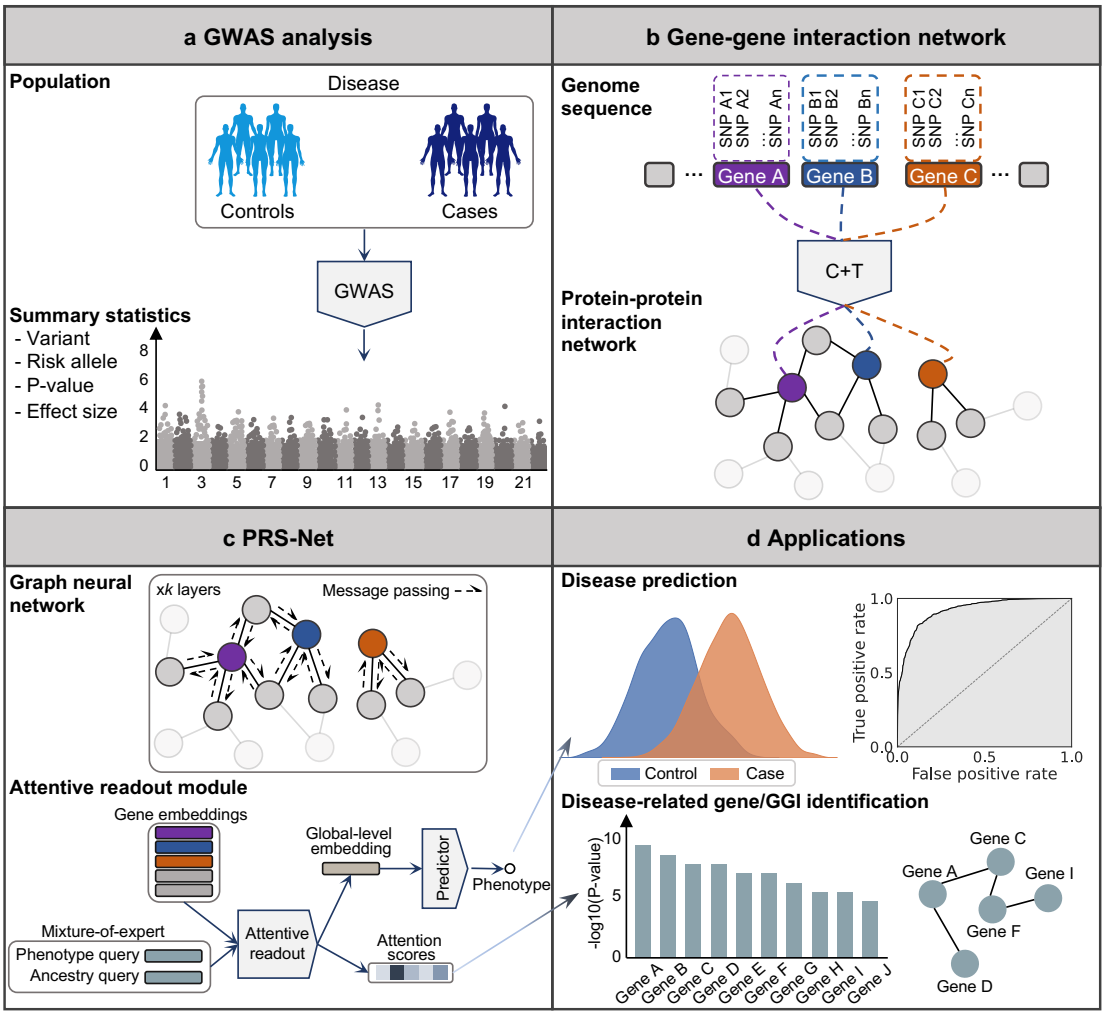

Fig. 1: An illustrative diagram of PRS-Net. **a** The proposed framework is based on summary statistics, including variants, risk alleles, P-values, and effect sizes derived from GWAS. **b** A gene-gene interaction network is constructed based on the protein-protein interaction network. Gene-level PRSs are calculated with the C+T method to serve as the node features for the nodes within the network. **c** A graph neural network is employed to update node features via message passing and subsequently an attentive readout module is applied to provide interpretable PRS predictions. **d** The PRS-Net can be applied for disease prediction and disease-related gene/GGI identification.

In this section, we present our proposed framework for PRS estimation (Fig. 1), covering the establishment of the GGI network, the derivation of gene-level PRS, and the architecture of PRS-Net.

2.1 GGI network

It is widely recognized that the disease phenotype is not solely determined by individual genes but rather involves the intricate interactions among multiple genes, which can exhibit additive or non-additive genetic relationships[35–37]. Additive genetic interactions manifest when the cumulative effects of genes jointly contribute to a specific phenotype. Furthermore, there are increasing studies highlighting the significance of non-additive genetic interactions[38–40]. Epistasis is a prominent example of non-additive genetic interaction, which occurs when the impact of a gene mutation depends on the presence or absence of mutations in one or more other genes[41–43]. We establish a GGI network that empowers PRS-Net to capture the intricate genetic relationships that are potentially associated with the target phenotypes (Fig. 1b).

We construct our GGI network based on the protein-protein interactions derived from the STRING database[44], as protein-protein interactions represent potent indicators of functional relationships between genes. Formally, we construct a GGI network, denoted as $\boldsymbol{\mathcal{G}} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}})$, where $\boldsymbol{\mathcal{V}}$ stands for the set of nodes and $\boldsymbol{\mathcal{E}}$ stands for the set of edges. Each node $v_i \in \boldsymbol{\mathcal{V}}$ stands for a coding gene and each edge $(v_i, v_j) \in \boldsymbol{\mathcal{E}}$ stands for an interaction between nodes v_i and v_j derived from the STRING database[44]. Note that, we add a self-loop (v_i, v_i) for each node $v_i \in \boldsymbol{\mathcal{V}}$. This network construction results in a GGI network encompassing 19,836 coding genes and 250,236 interactions.

Upon deriving the GGI network, we proceed to compute gene-level PRSs for the genes within the network using a C+T approach[11,21]. More precisely, for each gene in the network, we focus on the SNPs falling within a designated range, spanning from its transcription start site - L to its transcription end site + L. In our tests, we set L to 10 kilobases (KB), thereby encompassing the SNPs situated in non-coding regions, such as the promoters of the genes. Subsequently, for each gene, we perform LD clumping on the associated SNPs from the GWAS data, utilizing the LD information estimated in the target data. Following this, we filter the SNPs based on a specific P-value threshold. The gene-level PRSs are then derived by multiplying the genotype matrix by the effect sizes obtained from the GWAS data, and then dividing this by the number of allele observations for each gene. For the LD clumping process, we set the LD threshold R^2 to 0.5 and the physical distance threshold to 250 KB. As for the thresholding step, we set the P-values to $1e^{-5}$, $1e^{-4}$, $1e^{-3}$, $1e^{-2}$, $5e^{-2}$, 0.1, 0.2, 0.3, 0.5, and 1, respectively. This process results in the computation of eleven PRSs for each gene, which serves as their initial features. We denote the initial feature of $v_i \in \boldsymbol{\mathcal{V}}$ as $\boldsymbol{h}_i \in \boldsymbol{H}$, where $\boldsymbol{H} \in \mathbb{R}^{|\boldsymbol{\mathcal{V}}| \times 11}$ and $|\boldsymbol{\mathcal{V}}|$ stands for the number of genes in $\boldsymbol{\mathcal{G}}$.

2.2 PRS-Net

Graph neural network

We harness the power of a graph neural network to capture the complex interactions among genes within our established GGI network (Fig. 1c). In our tests, we specifically opt for a graph isomorphism network (GIN)[45] due to its proven theoretical and experimental expressiveness. Formally, we first encode the initial feature of nodes, denoted as $\boldsymbol{H}$, by employing an MLP in the following manner:

$$\boldsymbol{H}^0 = \mathrm{MLP}^0(\boldsymbol{H}), \tag{1}$$

where $\boldsymbol{H}^0 \in \mathbb{R}^{|\boldsymbol{\mathcal{V}}| \times D}$ and D is the dimension of hidden states. Subsequently, we apply multiple GIN layers to iteratively update the representation of each node by aggregating the representations of

its neighbors, as depicted below:

$$\boldsymbol{h}_i^k = \mathrm{MLP}^k((1+\epsilon^k)\cdot \boldsymbol{h}_i^{k-1} + \sum_{v_j \in \mathcal{N}(v_i)} \boldsymbol{h}_j^{k-1}), \tag{2}$$

where $\boldsymbol{h}_i^{k-1}$ is the hidden states of v_i at the $(k-1)$-th layer, $\mathcal{N}(v_i)$ stands for the neighbors of v_i in the GGI network, $\boldsymbol{h}_i^k$ stands for the updated hidden states of v_i at the k-th layer, MLP^k is the MLP at the k-th layer, and ϵ stands for a learnable variable. Following k iterations of aggregation, each gene effectively encapsulates the interaction information within its k-hop neighborhood.

Attentive readout module

To make predictions for each data sample, we derive the global-level representation for each sample through an attentive readout module, illustrated as follows:

$$\begin{aligned} \boldsymbol{h}_{\mathcal{G}} &= \text{Attentive readout}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}), \\ \boldsymbol{h}_{\mathcal{G}} &= \boldsymbol{A}\cdot \boldsymbol{V}, \\ \boldsymbol{A} &= \mathrm{Sigmoid}(\boldsymbol{Q}\cdot \boldsymbol{K}), \\ \boldsymbol{K} &= \boldsymbol{H}^k \cdot \boldsymbol{W}_K, \boldsymbol{V} = \boldsymbol{H}^k \cdot \boldsymbol{W}_V, \end{aligned} \tag{3}$$

where $\boldsymbol{W}_K \in \mathbb{R}^{D\times D}$ and $\boldsymbol{W}_V \in \mathbb{R}^{D\times D}$ stand for trainable projection matrices to derive the key (i.e., $\boldsymbol{K}$) and value (i.e., $\boldsymbol{V}$) matrices, respectively. $\boldsymbol{Q} \in \mathbb{R}^{1\times D}$ stands for a trainable query vector. $\mathrm{Sigmoid}$ stands for the sigmoid function. $\boldsymbol{A} \in \mathbb{R}^{1\times|\mathcal{V}|}$ stands for the attention scores, with elevated scores signifying a greater significance of the associated genes. $\boldsymbol{h}_{\mathcal{G}} \in \mathbb{R}^{1\times D}$ stands for the global-level representation.

After deriving the global-level representation, we employ an MLP to derive the final prediction, denoted as $\hat{\mathrm{PRS}}$, as follows:

$$\hat{\mathrm{PRS}} = \mathrm{MLP}(\boldsymbol{h}_{\mathcal{G}}). \tag{4}$$

Additionally, we implement a mixture-of-expert module[33] to effectively handle datasets that encompass data samples from multiple ancestries. More specifically, we introduce a specialized attentive readout module for each distinct ancestry. These dedicated attentive readout modules are activated when processing data from individuals with specific ancestral origins. To illustrate, when dealing with input samples of Western European ancestry, we derive the ancestry-specific global-level representation as follows:

$$\boldsymbol{h}_{\mathcal{G}}^{\mathrm{EUR}} = \text{Attentive readout}(\boldsymbol{Q}^{\mathrm{EUR}}, \boldsymbol{K}^{\mathrm{EUR}}, \boldsymbol{V}^{\mathrm{EUR}}). \tag{5}$$

The ancestry-specific readout module is designed to capture the unique knowledge pertaining to each ancestry in relation to the disease. In addition, we introduce another shared readout module to capture disease-related knowledge that holds general applicability across all ancestries:

$$\boldsymbol{h}_{\mathcal{G}}^{\mathrm{PH}} = \text{Attentive readout}(\boldsymbol{Q}^{\mathrm{PH}}, \boldsymbol{K}^{\mathrm{PH}}, \boldsymbol{V}^{\mathrm{PH}}). \tag{6}$$

Then, we derive the final global-level representation by combining the aforementioned two representations:

$$\boldsymbol{h}_{\mathcal{G}} = \boldsymbol{h}_{\mathcal{G}}^{\mathrm{EUR}} + \boldsymbol{h}_{\mathcal{G}}^{\mathrm{PH}}. \tag{7}$$

The process for deriving global-level representations of individuals from other ancestries follows a similar approach. The final PRS prediction can be computed with equation 4, utilizing the derived global-level representation. We refer to the single-ancestry variation as PRS-Net and the multiple-ancestry variation as PRS-Net$_{\mathrm{MA}}$.

3 Results

3.1 PRS-Net outperforms baseline methods in PRS prediction

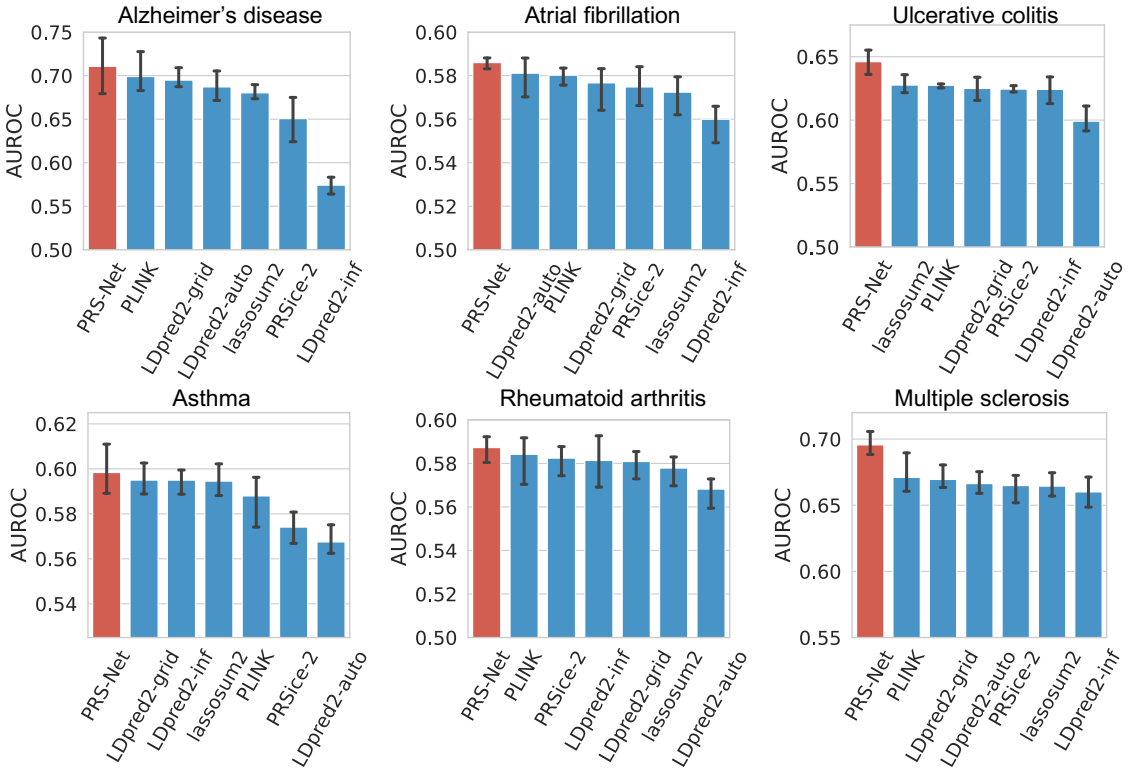

Fig. 2: The PRS prediction performance of PRS-Net compared to baseline methods across a range of complex diseases, including Alzheimer's disease, atrial fibrillation, ulcerative colitis, asthma, rheumatoid arthritis, and multiple sclerosis, measured in terms of the area under the receiver operating characteristic curve (AUROC). The bars are the estimated standard errors.

We extracted genotype-phenotype data from the UK Biobank database[34] for six different complex diseases, which encompassed Alzheimer's disease, atrial fibrillation, rheumatoid arthritis, multiple sclerosis, ulcerative colitis, and asthma. ICD-10 codes[46] were employed to define the disease phenotypes (Supplementary Table 1). For our primary experiments, we focused exclusively on individuals of Western European ancestry due to the insufficient size of the non-European ancestry population, which did not provide an adequate amount of training data (Supplementary Table 2). Following a quality control process, each disease dataset consisted of roughly 411,000 individuals (Supplementary Note 1.1). To prevent data leakage, we ensured that none of the GWAS were conducted on samples from the UK Biobank database (see Data availability). For each disease dataset, we randomly partitioned it into training, validation, and test sets with a ratio of 8:1:1. To evaluate the performance of PRS-Net, we compared it against several previously proposed methods, such as C+T-based methods (PLINK[21] and PRSice2[14]), lassosum2[17], and three variations of LDpred2[16] (LDpred2-auto, LDpred2-grid, and LDpred2-inf), utilizing the area under the receiver operating characteristic curve (AUROC) as the metric. To ensure a rigorous and equitable comparison, we utilized LD matrices estimated from European populations within the 1000

Genomes Project[47] across all methods in our study. Our results were based on three independent runs with different random seeds to ensure robustness and reliability. The results revealed that PRS-Net consistently outperformed all baseline methods on all disease datasets, resulting in relative improvements ranging from 0.5% to 3.7%. Interestingly, the largest improvements were obtained for two autoimmune diseases, i.e., ulcerative colitis (with a relative improvement of 3.0%) and multiple sclerosis (with a relative improvement of 3.7%), reinforcing the observed nonadditivity of genomic factors underlying these diseases[38,48–50]. Altogether, our data demonstrates that PRS-Net possesses the capacity to capture more intricate associations between genotypes and phenotypes that are beyond the reach of previously proposed linear models.

We utilized the Aalen-Johansen estimator[51] to estimate the disease occurrence over a lifetime for individuals categorized into high-risk and low-risk groups, as determined by the PRSs estimated by PRS-Net and baseline methods. High-risk individuals were defined as those with the highest 5% of PRSs, while low-risk individuals were identified as those with the lowest 5% of PRSs. The cumulative incidence plots revealed that individuals classified as high-risk by PRS-Net generally exhibited a heightened risk of disease throughout their lifetime compared to baseline methods, especially for ulcerative colitis, asthma, rheumatoid arthritis, and multiple sclerosis (Fig. 3a). Conversely, those categorized as low-risk by PRS-Net tended to maintain a lower risk of all diseases over their lifetime in comparison to baseline methods (Supplementary Fig. 1). These findings underscore the potential of PRS-Net as a powerful tool for individual risk stratification.

Next, we assessed the performance of PRS-Net and our multiple-ancestry model, PRS-Net$_{\mathrm{MA}}$, on a dataset comprising individuals from diverse ancestral backgrounds. Specifically, we curated a mixed-ancestry dataset encompassing Western European, South Asian, and African for asthma, which provides a reasonable number of asthma cases (over 1,000) for each ancestry (Supplementary Table 2). The results revealed that PRS-Net outperformed baseline methods on the mixed ancestry and South Asian ancestry test sets, indicating that the PRS-Net trained solely on the Western European ancestry dataset captured the underlying disease biology independent of different ancestries (Fig. 3b). Additionally, PRS-Net$_{\mathrm{MA}}$ demonstrated superior performance when compared to PRS-Net on the mixed ancestry, Western European ancestry, and African ancestry test sets (Fig. 3b). These findings underscored the ability of PRS-Net$_{\mathrm{MA}}$ to leverage the multi-ancestry dataset effectively, enhancing its portability in estimating PRS for individuals from diverse ancestral backgrounds.

3.2 PRS-Net identifies disease-related genes and GGIs for Alzheimer's disease and multiple sclerosis

Following the demonstration of the superior performance of PRS-Net in predicting PRS, we sought to explore its capability to identify risk genes and GGIs underlying complex diseases. Alzheimer's disease, a progressively degenerative condition, has been the subject of extensive research for many years, leading to the identification of numerous genes associated with the disease[52–56]. We employed PRS-Net to identify disease-related genes and GGIs, with the expectation that our findings would align with prior research outcomes. Specifically, we first applied the Mann–Whitney U test[57] to each gene within our constructed GGI network, assessing whether the attention scores associated with the gene for individuals with Alzheimer's disease were notably higher than those of the control group. This analysis yielded a gene set comprising 309 genes with compelling statistical significance (P-value $<$0.001). Please refer to Supplementary Data 1 for the complete list of the genes. Subsequently, we conducted gene set enrichment analyses (GSEA)[58] utilizing the gene ontology (GO)[59] and Kyoto Encyclopedia of Genes and Genomes (KEGG)[60] datasets on the identified gene set. Notably, the GO terms related to lipoprotein particles emerged as sig-

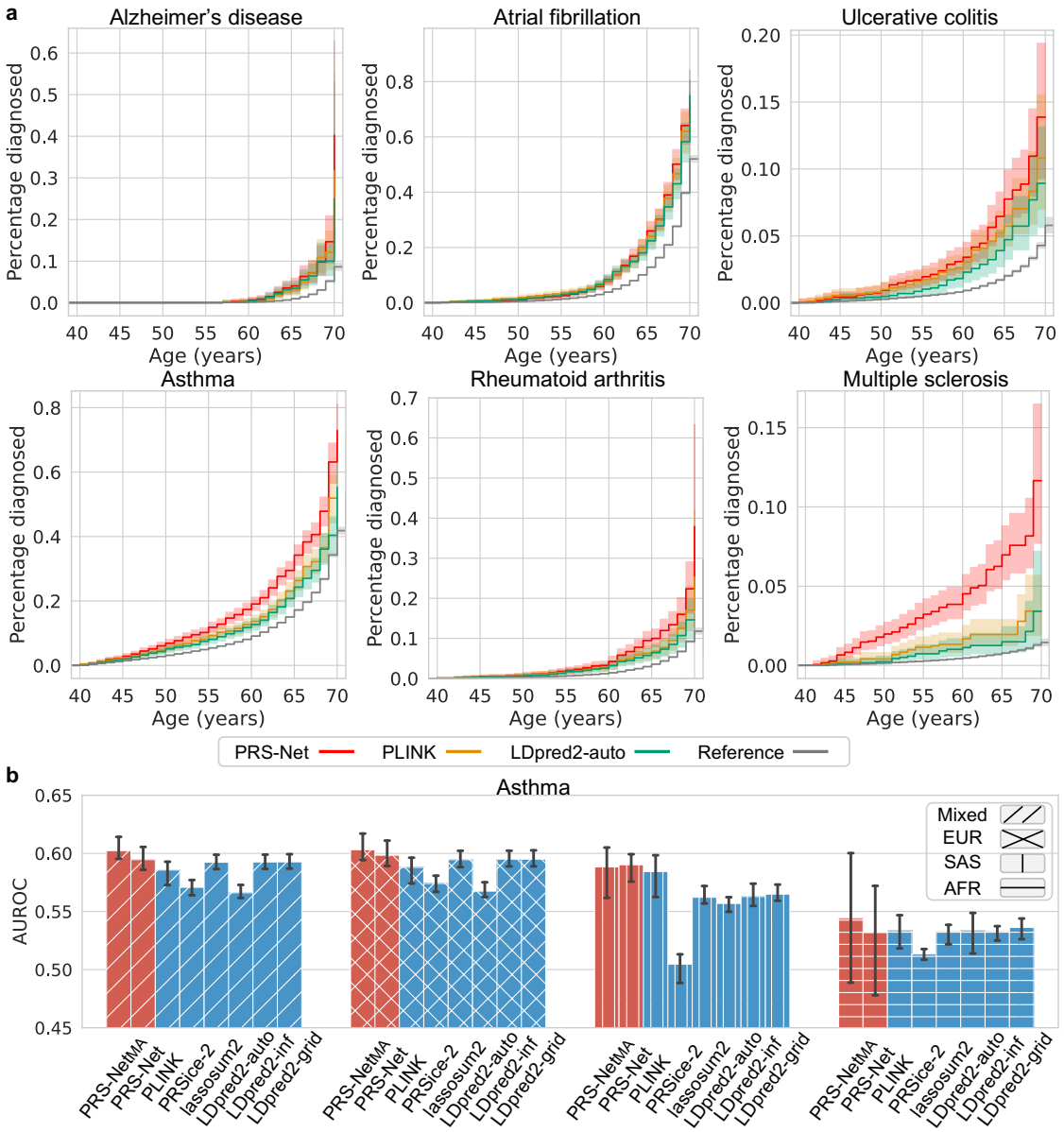

Fig. 3: **a** The cumulative incidence plots of high-risk individuals (with the highest 5% PRSs) identified by PRS-Net and baseline methods. Each plot illustrates the estimated percentage of individuals diagnosed with a specific disease at different ages. We provide cumulative incidence plots for the original datasets as a reference. **b** The PRS prediction performance of PRS-Net compared to baseline methods on an asthma dataset encompassing multiple ancestries, including Western European (EUR), South Asian (SAS), and African (AFR) ancestry, measured in terms of the area under the receiver operating characteristic curve (AUROC). The results on the mixed ancestry test set are also reported. The bars are the estimated standard errors.

nificantly enriched within the gene set (Supplementary Fig. 2a). This observation is in line with prior studies that have implicated lipoprotein particles as significantly potential risk factors for Alzheimer's diseasee[61–63] and have highlighted the role of metabolic dysregulation in the progression of Alzheimer's disease[64,65]. Notably, the exploration of high-density lipoprotein-inspired

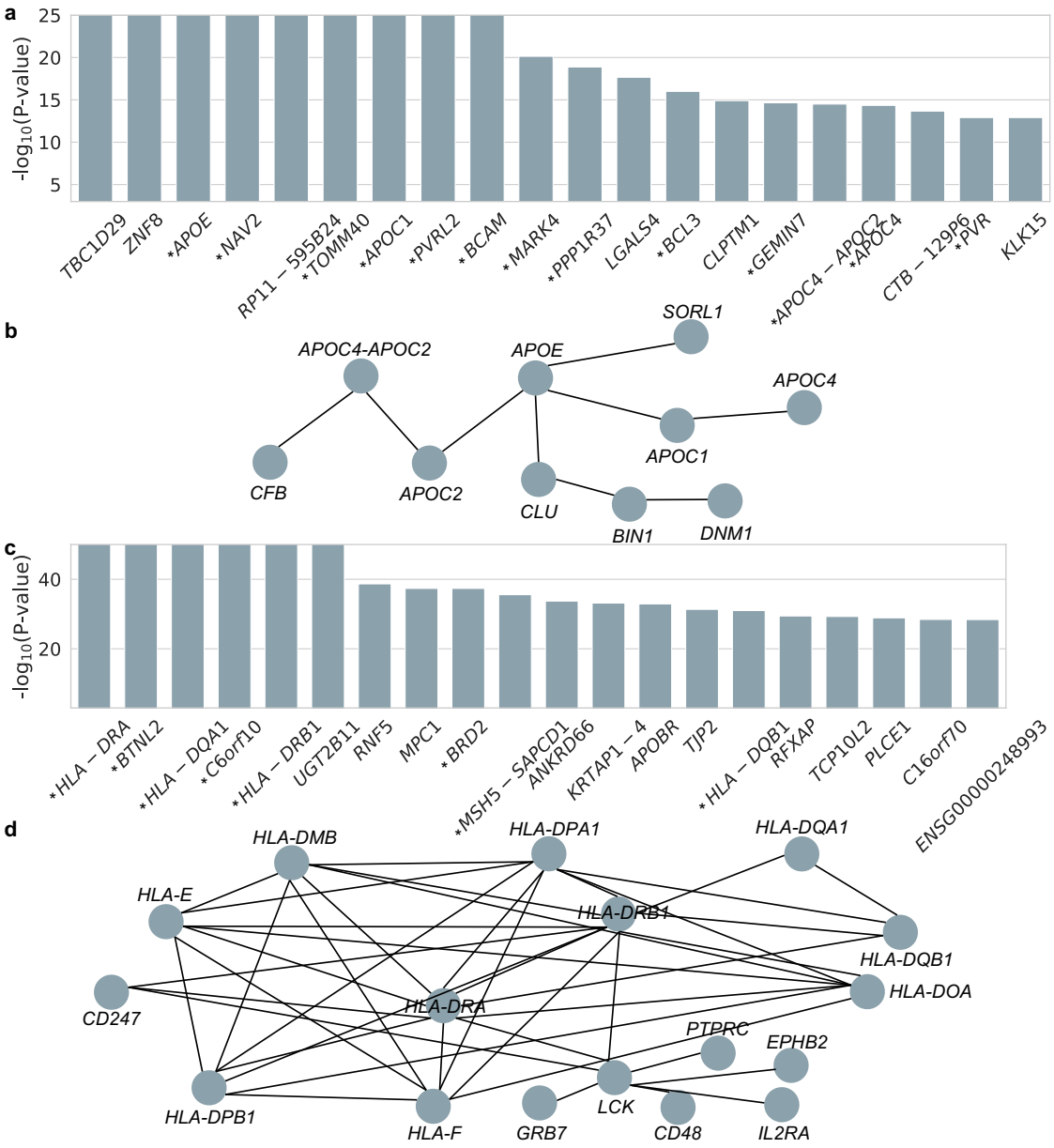

Fig. 4: PRS-Net identifies disease-related genes and GGIs for Alzheimer's disease and multiple sclerosis. **a** Top 20 genes with the highest statistical significance in the Mann-Whitney U test for Alzheimer's disease. The Mann–Whitney U test was utilized to assess whether the attention scores for a particular gene among the cases were significantly higher than those observed in the control group. An asterisk preceding the gene name signifies that the gene has been reported to be associated with Alzheimer's disease in previous studies. **b** Examples of interactions within the gene set with statistical significance (P-value <0.001) from the Mann-Whitney U test for Alzheimer's disease. **c** Top 20 genes with the highest statistical significance in the Mann-Whitney U test for multiple sclerosis. **d** Examples of interactions within the gene set with statistical significance (P-value <0.001) from the Mann-Whitney U test for multiple sclerosis.

221 treatments for Alzheimer's disease has been a well-documented area of study[62,63]. Fig. 4a il-

lustrates the top 20 genes with the utmost statistical significance in the Mann-Whitney U test. Remarkably, 15 out of these 20 genes have been identified as potential risk factors for Alzheimer's disease in previous studies. One notable example is *APOE*, which is the most prevalent high-density lipoprotein in the central nervous system and has been consistently linked to Alzheimer's disease in numerous studies[66–71]. Fig. 4b illustrates examples of the interactions from the GGI network between genes within the identified gene set. Please refer to Supplementary Data 2 for the complete list of the GGIs. Interestingly, aside from *APOE*, other genes within the *APOE* gene cluster, including *APOC1*, *APOC2*, and *APOC4*, were also identified as disease-related genes. This finding aligns with previous studies that have shown interdependent or independent associations of genes within the *APOE* gene cluster with Alzheimer's disease[72–76]. For instance, it has been shown that the variant *APOE* and *APOC2* exhibit interactive effects on metabolic pathways, potentially contributing to the risk of Alzheimer's disease[72]. *APOC1* also has been reported to serve as a risk factor for Alzheimer's disease in combination with *APOE*[74]. Furthermore, the combined effect of *APOE* and *CLU* on Alzheimer's disease has been observed[77]. *SORL1* is an APOE receptor gene, which has been recognized as a genetic risk factor in Alzheimer's disease. Recent research has elucidated the mechanistic connection between these two significant genetic factors in Alzheimer's disease[78]. A neuron-specific interaction between Alzheimer's disease risk factors *SORL1*, *APOE*, and *CLU* have also been shown in a recent study[79]. These observations highlight the proficiency of PRS-Net in not only identifying disease-related genes but also uncovering gene clusters that exhibit interactions contributing to the risk of Alzheimer's disease.

We also utilized PRS-Net to uncover genes and GGIs associated with multiple sclerosis. The Mann-Whitney U test identified a gene set with 456 potential risk genes (P-value <0.001). Please refer to Supplementary Data 3 for the complete list of the genes. The GSEA[58] using the KEGG[60] dataset on this gene set highlighted numerous immune-related pathways of statistical significance, such as antigen processing and presentation, allograft rejection, and graft-versus-host disease (Supplementary Fig. 4b). This finding aligns with the well-established understanding of multiple sclerosis as an autoimmune inflammatory disorder. The GSEA using the GO[59] dataset, unveiled significant enrichment of GO terms related to the major histocompatibility complex (MHC) protein complex within the identified gene set (Supplementary Fig. 4a), which can be supported by previous studies that underscore the substantial genetic impact of MHCs on multiple sclerosis[80–84]. *HLA-DRA*, a subunit of *HLA-DR* which is a human MHC, was identified as the most significant gene in our analysis (Fig. 4c). Moreover, substantial *HLA* genes were identified as risk genes in our analysis (Fig. 4d). Please refer to Supplementary Data 4 for the complete list of the GGIs. This finding is in line with a previous study indicating that *HLA* interactions modulate genetic risk for multiple sclerosis[85]. Additionally, non-additive interactions between *HLA*s have been widely reported to significantly affect the risk of autoimmune diseases[38,48–50]. These discoveries collectively provide compelling evidence of the potential of PRS-Net to offer valuable insights that advance our understanding of diseases.

3.3 Ablation studies

To assess the effectiveness of specific design choices in PRS-Net, we conducted comprehensive ablation studies. We introduced various modified frameworks derived from PRS-Net, each with distinct constraints: PRS-Net-GGI (omitting the GGI network), PRS-Net-Att+Sum (replacing the attentive readout module with a sum readout module, which summarized the node feature to derive the global-level representations), PRS-Net-Att+Mean (replacing the attentive readout module with a mean readout module, which computes the average of node features to derive global-level representations), and PRS-Net-Att+Max (replacing the attentive readout module with a max read-

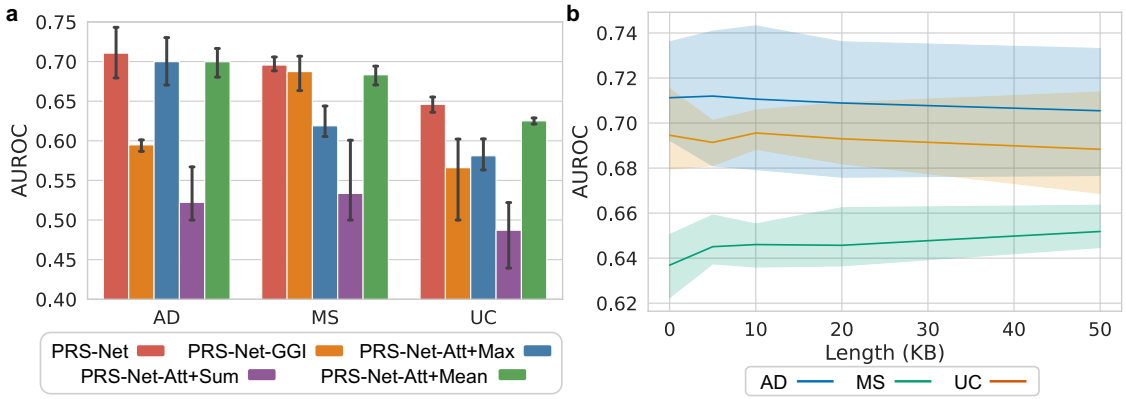

Fig. 5: The results of ablation studies on PRS-Net. **a** The comparison results of PRS-Net and its variations, including PRS-Net-GGI (omitting the GGI network), PRS-Net-Att+Sum (replacing the attentive readout module with a sum readout module, which summarized the node feature to derive the global-level representations), PRS-Net-Att+Mean (replacing the attentive readout module with a mean readout module, which computes the average of node features to derive global-level representations), and PRS-Net-Att+Max (replacing the attentive readout module with a max readout module, which extracts maximum values across node features to derive the global-level representations), conducted on the datasets of Alzheimer's disease (AD), multiple sclerosis (MS), and ulcerative colitis (UC). The bars are the estimated standard errors. **b** The PRS prediction performance of PRS-Net versus the extension lengths upstream and downstream of the transcription start and end sites.

out module, which extracts maximum values across node features to derive the global-level representations). We compared the performance of PRS-Net against these variants using datasets related to Alzheimer's disease, multiple sclerosis, and ulcerative colitis. The results showcased that PRS-Net surpassed PRS-Net-GGI by an average relative improvement of 11.6%, underscoring the significance of incorporating the GGI network to capture the intricate genetic interactions associated with diseases (Fig. 5a). Furthermore, PRS-Net outperformed PRS-Net-Att+Sum, PRS-Net-Att+Mean, and PRS-Net-Att+Max with average relative improvements of 33.0%, 2.2%, and 8.4%, respectively, highlighting the effectiveness of the attentive readout module in summarizing node features (Fig. 5a).

Additionally, we explored the impact of varying extension lengths both upstream and downstream of the transcription start and end sites when calculating gene-level PRSs. We assessed different length values, including 0, 5, 10, 20, and 50 KB, and subsequently evaluated their prediction performance. The results demonstrated that PRS-Net is generally robust to different extension lengths (Fig. 5b). However, it is noteworthy that the performance of PRS-Net on the multiple sclerosis dataset significantly declined when the extension length was set to 0 KB (Fig. 5b). This observation suggested that including SNPs from non-coding regions can indeed enhance the accuracy of PRS prediction.

Discussion

In this study, we develop PRS-Net, a deep-learning framework that offers interpretable and improved PRS predictions. By constructing a GGI network and incorporating a graph neural network, PRS-Net fully takes advantage of the power of non-linear associations between genetic

factors and phenotypes. Additionally, the integration of an attentive readout module empowers PRS-Net to deliver interpretable predictions. Through comprehensive testing across six complex diseases, PRS-Net consistently achieved superior performance in comparison with baseline methods in PRS prediction. Furthermore, we demonstrated the interpretability of PRS-Net by using it to identify specific genes and GGIs that significantly impact the risk of Alzheimer's disease and multiple sclerosis. In summary, PRS-Net provides a potent tool for accurate PRS prediction and biological discovery for complex diseases.

Data availability

The GWAS data for Alzheimer's disease can be accessed at https://ctg.cncr.nl/software/summary_statistics/. The GWAS data for atrial fibrillation can be accessed at https://cvd.hugeamp.org/downloads.html#summary/. The GWAS data for ulcerative colitis can be accessed at ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2016-11-07/. The GWAS data for asthma can be accessed at https://www.globalbiobankmeta.org/resources/. The GWAS data for rheumatoid arthritis can be accessed at https://data.cyverse.org/dav-anon/iplant/home/kazuyoshiishigaki/ra_gwas/ra_gwas-10-28-2021.tar/. The GWAS data for multiple sclerosis can be accessed at https://imsgc.net/?page_id=31/. The UKBB dataset is available at https://www.ukbiobank.ac.uk.

Code availability

The source code of PRS-Net can be downloaded from the Github repository at https://github.com/lihan97/PRS-Net.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (T2125007 to J.Z.), the National Key Research and Development Program of China (2021YFF1201300 to J.Z.), the New Cornerstone Science Foundation through the XPLORER PRIZE (J.Z.), the Research Center for Industries of the Future (RCIF) at Westlake University (J.Z.), and the Westlake Education Foundation (J.Z.).

Author contributions statement

H.L. and S.Z. conceived the concept and designed the study. H.L. and S.Z. developed the methodology and conducted data analysis. H.L., J.Z., M.S. and S.Z. are responsible for the data interpretation. S.Z., M.S. and J.Z. supervised the project. H.L. and S.Z. prepared the manuscript with the assistance from all other authors.

Competing interests statement

All authors declare no competing interests.

References

1. Robert Clarke, John F Peden, Jemma C Hopewell, Theodosios Kyriakou, Anuj Goel, Simon C Heath, Sarah Parish, Simona Barlera, Maria Grazia Franzosi, Stephan Rust, et al. Genetic variants associated with lp (a) lipoprotein level and coronary disease. New England Journal of Medicine, 361(26):2518–2528, 2009.
2. CKDGen Consortium, KidneyGen Consortium, EchoGen consortium, CHARGE-HF consortium, Thor Aspelund, Melissa Garcia, Yen-Pei C Chang, Jeffrey R O'Connell, Nanette I Steinle, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature, 478(7367):103–109, 2011.
3. Chris Gaiteri, Sara Mostafavi, Christopher J Honey, Philip L De Jager, and David A Bennett. Genetic variants in alzheimer disease—molecular and brain network approaches. Nature Reviews Neurology, 12(7):413–427, 2016.
4. Valeria Orrù, Maristella Steri, Gabriella Sole, Carlo Sidore, Francesca Virdis, Mariano Dei, Sandra Lai, Magdalena Zoledziewska, Fabio Busonero, Antonella Mulas, et al. Genetic variants regulating immune cell levels in health and disease. Cell, 155(1):242–256, 2013.
5. Ali Torkamani, Nathan E Wineinger, and Eric J Topol. The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics, 19(9):581–590, 2018.
6. Cathryn M Lewis and Evangelos Vassos. Polygenic risk scores: from research tools to clinical instruments. Genome medicine, 12(1):1–11, 2020.
7. Shing Wan Choi, Timothy Shin-Heng Mak, and Paul F O'Reilly. Tutorial: a guide to performing polygenic risk score analyses. Nature protocols, 15(9):2759–2772, 2020.
8. Greg Gibson. On the utilization of polygenic risk scores for therapeutic targeting. PLoS genetics, 15(4):e1008060, 2019.
9. Takahiro Konuma and Yukinori Okada. Statistical genetics and polygenic risk score for precision medicine. Inflammation and regeneration, 41(1):1–5, 2021.
10. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. Nature Medicine, 27(11):1876–1884, 2021.
11. Bjarni J Vilhjálmsson, Jian Yang, Hilary K Finucane, Alexander Gusev, Sara Lindström, Stephan Ripke, Giulio Genovese, Po-Ru Loh, Gaurav Bhatia, Ron Do, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. The american journal of human genetics, 97(4):576–592, 2015.
12. Jack Euesden, Cathryn M Lewis, and Paul F O'reilly. Prsice: polygenic risk score software. Bioinformatics, 31(9):1466–1468, 2015.
13. Timothy Shin Heng Mak, Robert Milan Porsch, Shing Wan Choi, Xueya Zhou, and Pak Chung Sham. Polygenic scores via penalized regression on summary statistics. Genetic epidemiology, 41(6):469–480, 2017.
14. Shing Wan Choi and Paul F O'Reilly. Prsice-2: Polygenic risk score software for biobank-scale data. Gigascience, 8(7):giz082, 2019.
15. Luke R Lloyd-Jones, Jian Zeng, Julia Sidorenko, Loïc Yengo, Gerhard Moser, Kathryn E Kemper, Huanwei Wang, Zhili Zheng, Reedik Magi, Tõnu Esko, et al. Improved polygenic prediction by bayesian multiple regression on summary statistics. Nature communications, 10(1):5086, 2019.
16. Florian Privé, Julyan Arbel, and Bjarni J Vilhjálmsson. Ldpred2: better, faster, stronger. Bioinformatics, 36(22-23):5424–5431, 2020.
17. Florian Privé, Julyan Arbel, Hugues Aschard, and Bjarni J Vilhjálmsson. Identifying and correcting for misspecifications in gwas summary statistics and polygenic scores. Human Genetics and Genomics Advances, 3(4), 2022.
18. William YS Wang, Bryan J Barratt, David G Clayton, and John A Todd. Genome-wide association studies: theoretical and practical concerns. Nature Reviews Genetics, 6(2):109–118, 2005.
19. Arthur Korte and Ashley Farlow. The advantages and limitations of trait analysis with gwas: a review. Plant methods, 9(1):1–9, 2013.
20. Emil Uffelmann, Qin Qin Huang, Nchangwi Syntia Munung, Jantina De Vries, Yukinori Okada, Alicia R Martin, Hilary C Martin, Tuuli Lappalainen, and Danielle Posthuma. Genome-wide association studies. Nature Reviews Methods Primers, 1(1):59, 2021.
21. Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel AR Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul IW De Bakker, Mark J Daly, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics, 81(3):559–575, 2007.
22. Ying Wang, Kristin Tsuo, Masahiro Kanai, Benjamin M Neale, and Alicia R Martin. Challenges and opportunities for developing more generalizable polygenic risk scores. Annual review of biomedical data science, 5:293–320, 2022.
23. Tin Kam Ho. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition, volume 1, pages 278–282. IEEE, 1995.
24. Leo Breiman. Random forests. Machine learning, 45:5–32, 2001.
25. Jerome H Friedman. Greedy function approximation: a gradient boosting machine. Annals of statistics, pages 1189–1232, 2001.

26. Jerome H Friedman. Stochastic gradient boosting. Computational statistics & data analysis, 38(4):367–378, 2002.
27. Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1):119–139, 1997.
28. Trevor Hastie, Saharon Rosset, Ji Zhu, and Hui Zou. Multi-class adaboost. Statistics and its Interface, 2(3):349–360, 2009.
29. Hadeel Alzoubi, Raid Alzubi, and Naeem Ramzan. Deep learning framework for complex disease risk prediction using genomic variations. Sensors, 23(9):4439, 2023.
30. Arnór I Sigurdsson, Ioannis Louloudis, Karina Banasik, David Westergaard, Ole Winther, Ole Lund, Sisse Rye Ostrowski, Christian Erikstrup, Ole Birger Vesterager Pedersen, Mette Nyegaard, et al. Deep integrative models for large-scale human genomics. Nucleic Acids Research, 51(12):e67–e67, 2023.
31. Pau Bellot, Gustavo de Los Campos, and Miguel Pérez-Enciso. Can deep learning improve genomic prediction of complex human traits? Genetics, 210(3):809–819, 2018.
32. Yu Xu, Dragana Vuckovic, Scott C Ritchie, Parsa Akbari, Tao Jiang, Jason Grealey, Adam S Butterworth, Willem H Ouwehand, David J Roberts, Emanuele Di Angelantonio, et al. Machine learning optimized polygenic scores for blood cell traits identify sex-specific trajectories and genetic correlations with disease. Cell Genomics, 2(1), 2022.
33. Saeed Masoudnia and Reza Ebrahimpour. Mixture of experts: a literature survey. Artificial Intelligence Review, 42:275–293, 2014.
34. Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine, 12(3):e1001779, 2015.
35. Heather J Cordell. Detecting gene–gene interactions that underlie human diseases. Nature Reviews Genetics, 10(6):392–404, 2009.
36. Or Zuk, Eliana Hechter, Shamil R Sunyaev, and Eric S Lander. The mystery of missing heritability: Genetic interactions create phantom heritability. Proceedings of the National Academy of Sciences, 109(4):1193–1198, 2012.
37. Matthew B Taylor and Ian M Ehrenreich. Higher-order genetic interactions and their contribution to complex traits. Trends in genetics, 31(1):34–40, 2015.
38. Tobias L Lenz, Aaron J Deutsch, Buhm Han, Xinli Hu, Yukinori Okada, Stephen Eyre, Michael Knapp, Alexandra Zhernakova, Tom WJ Huizinga, Goncalo Abecasis, et al. Widespread non-additive and interaction effects within hla loci modulate the risk of autoimmune diseases. Nature genetics, 47(9):1085–1090, 2015.
39. Luis Varona, Andres Legarra, Miguel A Toro, and Zulma G Vitezica. Non-additive effects in genomic selection. Frontiers in genetics, 9:78, 2018.
40. Marta Guindo-Martínez, Ramon Amela, Silvia Bonàs-Guarch, Montserrat Puiggròs, Cecilia Salvoro, Irene Miguel-Escalada, Caitlin E Carey, Joanne B Cole, Sina Rüeger, Elizabeth Atkinson, et al. The impact of non-additive genetic associations on age-related complex diseases. Nature communications, 12(1):2436, 2021.
41. James M Cheverud and Eric J Routman. Epistasis and its contribution to genetic variance components. Genetics, 139(3):1455–1461, 1995.
42. Jason H Moore and Scott M Williams. Epistasis and its implications for personal genetics. The American Journal of Human Genetics, 85(3):309–320, 2009.
43. Ben Lehner. Molecular mechanisms of epistasis within and between genes. Trends in Genetics, 27(8):323–331, 2011.
44. Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L Gable, Tao Fang, Nadezhda T Doncheva, Sampo Pyysalo, et al. The string database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic acids research, 51(D1):D638–D646, 2023.
45. Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In International Conference on Learning Representations, 2018.
46. World Health Organization et al. Icd-10. international statistical classification of diseases and related health problems: Tenth revision 1992, volume 1= cim-10. classification statistique internationale des maladies et des problèmes de santé connexes: Dixième révision 1992, volume 1. 1992.
47. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature, 526(7571):68, 2015.
48. Marc Lipsitch, Carl T Bergstrom, and Rustom Antia. Effect of human leukocyte antigen heterozygosity on infectious disease outcome: the need for allele-specific measures. BMC Medical Genetics, 4(1):1–9, 2003.
49. Sue Tsai and Pere Santamaria. Mhc class ii polymorphisms, autoreactive t-cells, and autoimmunity. Frontiers in immunology, 4:321, 2013.
50. Philippe Goyette, Gabrielle Boucher, Dermot Mallon, Eva Ellinghaus, Luke Jostins, Hailiang Huang, Stephan Ripke, Elena S Gusareva, Vito Annese, Stephen L Hauser, et al. High-density mapping of the mhc identifies a shared role for hla-drb1* 01: 03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nature genetics, 47(2):172–179, 2015.

51. Odd O Aalen and Søren Johansen. An empirical transition matrix for non-homogeneous markov chains based on censored observations. Scandinavian Journal of Statistics, pages 141–150, 1978.
52. Dennis J Selkoe. Alzheimer's disease: genes, proteins, and therapy. Physiological reviews, 2001.
53. Rudolph E Tanzi and Lars Bertram. New frontiers in alzheimer's disease genetics. Neuron, 32(2):181–184, 2001.
54. Lars Bertram and Rudolph E Tanzi. Thirty years of alzheimer's disease genetics: the implications of systematic meta-analyses. Nature Reviews Neuroscience, 9(10):768–778, 2008.
55. Celeste M Karch, Carlos Cruchaga, and Alison M Goate. Alzheimer's disease genetics: from the bench to the clinic. Neuron, 83(1):11–26, 2014.
56. Ya-Ping Tang and Elliot S Gershon. Genetic studies in alzheimer's disease. Dialogues in clinical neuroscience, 2022.
57. Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, pages 50–60, 1947.
58. Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102(43):15545–15550, 2005.
59. Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. Nature genetics, 25(1):25–29, 2000.
60. Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1):27–30, 2000.
61. Zhike Zhou, Yifan Liang, Xiaoqian Zhang, Junjie Xu, Jueying Lin, Rongwei Zhang, Kexin Kang, Chang Liu, Chuansheng Zhao, and Mei Zhao. Low-density lipoprotein cholesterol and alzheimer's disease: a systematic review and meta-analysis. Frontiers in aging neuroscience, 12:5, 2020.
62. Mirna Safieh, Amos D Korczyn, and Daniel M Michaelson. Apoe4: an emerging therapeutic target for alzheimer's disease. BMC medicine, 17:1–17, 2019.
63. Yi Jin, Kudzai Chifodya, Guochen Han, Wenxin Jiang, Yun Chen, Yang Shi, Qiao Xu, Yilong Xi, Jun Wang, Jianping Zhou, et al. High-density lipoprotein in alzheimer's disease: From potential biomarkers to therapeutics. Journal of Controlled Release, 338:56–70, 2021.
64. Suzanne Craft. The role of metabolic disorders in alzheimer disease and vascular dementia: two roads converged. Archives of neurology, 66(3):300–305, 2009.
65. Xu Yan, Yue Hu, Biyao Wang, Sijian Wang, and Xinwen Zhang. Metabolic dysregulation contributes to the progression of alzheimer's disease. Frontiers in neuroscience, 14:530219, 2020.
66. MS Tsai, Eric G Tangalos, Ronald C Petersen, Glenn E Smith, Daniel J Schaid, Emre Kokmen, Robert J Ivnik, and Stephen N Thibodeau. Apolipoprotein e: risk factor for alzheimer disease. American journal of human genetics, 54(4):643, 1994.
67. Warren J Strittmatter and Allen D Roses. Apolipoprotein e and alzheimer disease. Proceedings of the National Academy of Sciences, 92(11):4725–4727, 1995.
68. Marion R Meyer, JoAnn T Tschanz, Maria C Norton, Kathleen A Welsh-Bohmer, David C Steffens, Bonita W Wyse, and John Breitner. Apoe genotype predicts when—not whether—one is predisposed to develop alzheimer disease. Nature genetics, 19(4):321–322, 1998.
69. Robert C Green, J Scott Roberts, L Adrienne Cupples, Norman R Relkin, Peter J Whitehouse, Tamsen Brown, Susan LaRusse Eckert, Melissa Butson, A Dessa Sadovnick, Kimberly A Quaid, et al. Disclosure of apoe genotype for risk of alzheimer's disease. New England Journal of Medicine, 361(3):245–254, 2009.
70. Emmanuelle Genin, Didier Hannequin, David Wallon, Kristel Sleegers, Mikko Hiltunen, Onofre Combarros, Maria Jesús Bullido, Sebastiaan Engelborghs, Peter De Deyn, Claudine Berr, et al. Apoe and alzheimer disease: a major gene with semi-dominant inheritance. Molecular psychiatry, 16(9):903–907, 2011.
71. Yu Yamazaki, Na Zhao, Thomas R Caulfield, Chia-Chen Liu, and Guojun Bu. Apolipoprotein e and alzheimer disease: pathobiology and targeting strategies. Nature Reviews Neurology, 15(9):501–518, 2019.
72. Robert A Hegele, W Carl Breckenridge, Diane W Cox, Graham F Maguire, J Alick Little, and Philip W Connelly. Interaction between variant apolipoproteins c-ii and e that affects plasma lipoprotein concentrations. Arteriosclerosis and Thrombosis: A Journal of Vascular Biology, 11(5):1303–1309, 1991.
73. Sebastián Cervantes, Lluís Samaranch, José Manuel Vidal-Taboada, Isabel Lamet, María Jesús Bullido, Ana Frank-García, Francisco Coria, Albert Lleó, Jordi Clarimón, Elena Lorenzo, et al. Genetic variation in apoe cluster region and alzheimer's disease risk. Neurobiology of aging, 32(11):2107–e7, 2011.
74. Qin Zhou, Fan Zhao, Ze-ping Lv, Chen-guang Zheng, Wei-dong Zheng, Liang Sun, Na-na Wang, Shenghang Pang, Fabiana Michelsen de Andrade, Mian Fu, et al. Association between apoc1 polymorphism and alzheimer's disease: a case-control study and meta-analysis. PloS one, 9(1):e87017, 2014.

75. Yvonne Shao, McKenzie Shaw, Kaitlin Todd, Maria Khrestian, Giana D'Aleo, P John Barnard, Jeff Zahratka, Jagan Pillai, Chang-En Yu, C Dirk Keene, et al. Dna methylation of tomm40-apoe-apoc2 in alzheimer's disease. Journal of human genetics, 63(4):459–471, 2018.
76. Alexander M Kulminski, Ethan Jain-Washburn, Ian Philipp, Liang He, Yury Loika, Elena Loiko, Olivia Bagley, Svetlana Ukraintseva, Anatoliy Yashin, Konstantin Arbeev, et al. Apoe epsilon4 allele and tomm40-apoc1 variants jointly contribute to survival to older ages. Aging Cell, 21(12):e13730, 2022.
77. Florence F Roussotte, Boris A Gutman, Sarah K Madsen, John B Colby, Paul M Thompson, Alzheimer's Disease Neuroimaging Initiative, et al. Combined effects of alzheimer risk variants in the clu and apoe genes on ventricular expansion patterns in the elderly. Journal of Neuroscience, 34(19):6537–6545, 2014.
78. Hyo Lee, Aimee J Aylward, Richard V Pearse, Yi-Chen Hsieh, Zachary M Augur, Courtney R Benoit, Vicky Chou, Allison Knupp, Cheryl Pan, Srilakshmi Goberdhan, et al. Cell-type-specific regulation of apoe levels in human neurons by the alzheimer's disease risk gene sorl1. bioRxiv, pages 2023–02, 2023.
79. Pranav Preman and Amaia M Arranz. A neuron-specific interaction between alzheimer's disease risk factors sorl1, apoe, and clu. Cell Reports, 42(9), 2023.
80. Ute Traugott. Multiple sclerosis: relevance of class i and class ii mhc-expressing cells to lesion development. Journal of neuroimmunology, 16(2):283–302, 1987.
81. Sunhee C Lee, GR Wayne Moore, George Golenwsky, and Cedric S Raine. Multiple sclerosis: a role for astroglia in active demyelination suggested by class ii mhc expression and ultrastructural study. Journal of neuropathology and experimental neurology, 49(2):122–136, 1990.
82. Multiple Sclerosis Genetics Group, Jonathan L Haines, Henry A Terwedow, Katie Burgess, Margaret A Pericak-Vance, Jackie B Rimmler, Eden R Martin, Jorge R Oksenberg, Robin Lincoln, David Y Zhang, et al. Linkage of the mhc to familial multiple sclerosis suggests genetic heterogeneity. Human molecular genetics, 7(8):1229–1234, 1998.
83. Matthew R Lincoln, Alexandre Montpetit, M Zameel Cader, Janna Saarela, David A Dyment, Milvi Tiislar, Vincent Ferretti, Pentti J Tienari, A Dessa Sadovnick, Leena Peltonen, et al. A predominant role for the hla class ii region in the association of the mhc region with multiple sclerosis. Nature genetics, 37(10):1108–1112, 2005.
84. David A Dyment, Blanca M Herrera, M Zameel Cader, Cristen J Willer, Matthew R Lincoln, A Dessa Sadovnick, Neil Risch, and George C Ebers. Complex interactions among mhc haplotypes in multiple sclerosis: susceptibility and resistance. Human molecular genetics, 14(14):2019–2026, 2005.
85. Class ii hla interactions modulate genetic risk for multiple sclerosis. Nature genetics, 47(10):1107–1113, 2015.