Integrating GWAS and Transcriptome Data through PrediXcan and Multimodal Deep Learning to Reveal Genetic Basis and Novel Drug Repositioning Opportunities for Alzheimer's Disease

Xuecong Tian[a, #], Ying Su[a, #], Sizhe Zhang [b,#], Haiqing Sun[b], Chen Chen[b], Cheng Chen[b], Luyu Zhou[c] , Xiaoyi Lv[b,*], Panos Roussos [d,*], Wen Zhang[e,*]

[#] These authors contributed equally to this work

[a] School of Computer Science and Technology, Xinjiang University, Urumqi 830046, Xinjiang, China;

[b] College of Software, Xinjiang University, Urumqi 830046, Xinjiang, China;

[c] Department of Pharmacy, College of Biology, Hunan University, Changsha, Hunan 410082, China;

[d] Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA;

[e] Charles E. Schmidt College of Science, Florida Atlantic University, Boca Raton, FL 33431, USA;

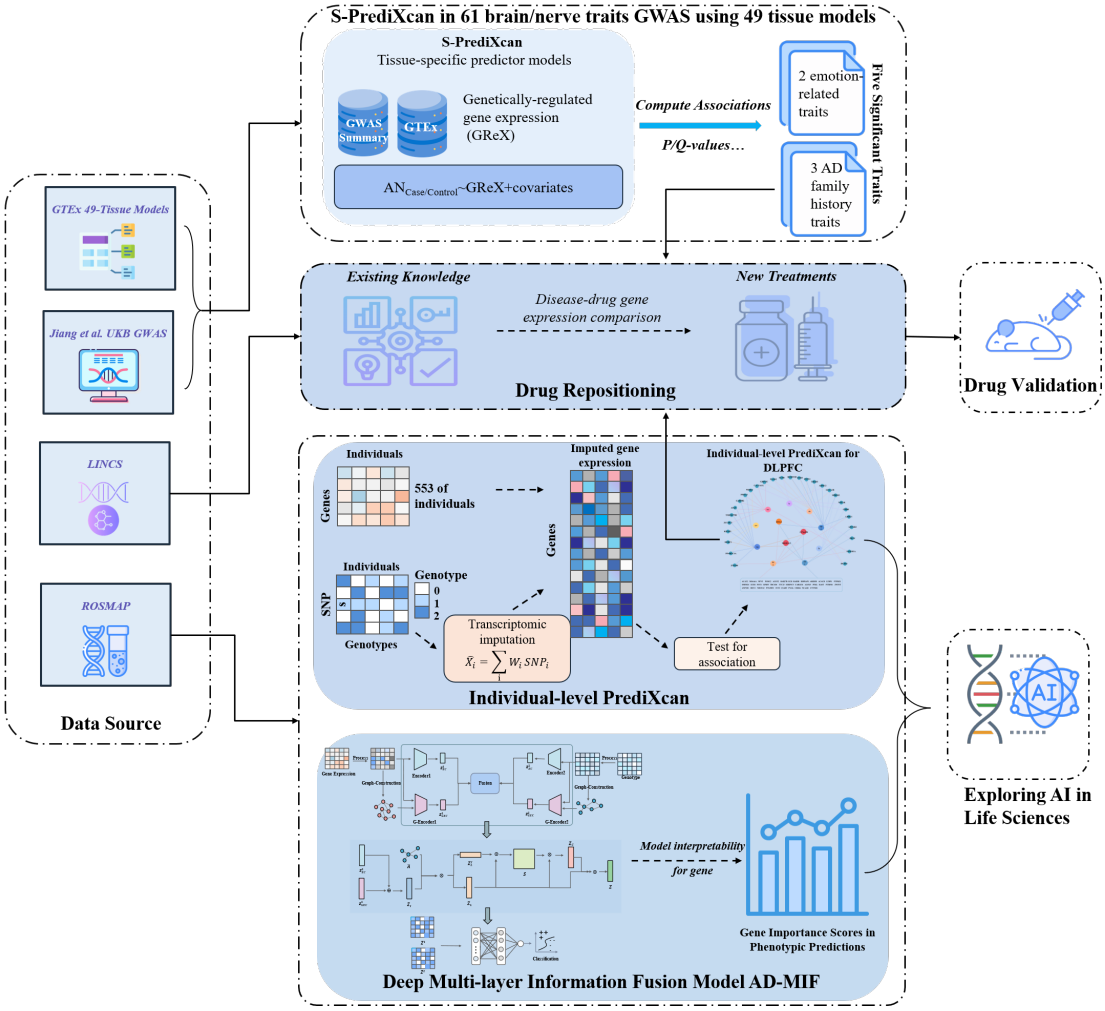
* Corresponding authors. E-mail addresses: xjuwawj01@163.com (X. Lv); panagiotis.roussos@mssm.edu (P. Roussos); zhangw@fau.edu (W. Zhang).

Abstract

Alzheimer's Disease (AD) is the leading cause of dementia, imposing significant economic and social burdens. Although genome-wide association studies (GWAS) have identified approximately 70 risk loci, the functional mechanisms underlying AD remain unclear. In this study, we integrated the GWAS summary statistics from Jiang et al. with gene expression data from the GTEx project using S-PrediXcan method, encompassing 61 brain-related traits across 49 tissues. Comprehensive analysis identified five significant traits, including family history of AD, and highlighted key genes such as APOE, APOC1, and TOMM40, which play crucial roles in cholesterol metabolism, immune response, and neuroinflammation. Validation using the ROSMAP dataset confirmed the association of these genes with AD phenotypes. Furthermore, we developed AD-MIF, a novel deep multi-layer information fusion model that integrates multi-omics data, achieving a 10-20% improvement in AUC performance for predicting AD-related traits compared to traditional models. Gene enrichment analysis emphasized the importance of pathways such as cholesterol metabolism and immune

response in the pathogenesis of AD. Additionally, drug repositioning analysis identified candidate drugs, including Dasatinib and Sirolimus, which may alleviate AD progression by reducing neuroinflammation and clearing senescent cells. Our findings advance the understanding of the genetic architecture of AD, improve predictive models, and propose potential therapeutic drugs.

Graphical Abstract



Main

Alzheimer's Disease (AD) is the leading cause of dementia and is rapidly becoming one of the most expensive, fatal, and burdensome diseases of this century [1, 2]. This neurodegenerative disease impairs cognitive functions such as memory, behavior, and emotions, ultimately leading to the loss of patients' ability to care for themselves and engage socially [3]. AD primarily affects the cerebral cortex and hippocampal regions, with its pathological features associated with insoluble amyloid-beta (Aβ) plaques and tau protein aggregates within neurons [4]. Currently, approximately 6.9 million Americans aged 65 and older are affected by AD, and medical and caregiving costs are expected to reach $360 billion by 2024 [5]. To alleviate the heavy burden posed by AD, it is urgently necessary to explore its etiology and identify potential therapeutic drugs.

Genome-Wide Association Studies (GWAS) are experimental designs used to detect associations between genetic variations and traits, with results depending on factors such as sample size and variant frequency [6]. To date, GWAS has identified approximately 70 genetic loci associated with AD risk, providing important clues for studying the genetic mechanisms of AD [7]. However, GWAS loci are difficult to interpret, which cannot precisely identify functional genes and regulatory mechanisms [8]. Transcriptome-Wide Association Studies (TWAS) integrate GWAS data with gene expression changes, using reference gene expression data to predict the expression levels of disease-related genes, thereby identifying genes associated with traits on a genome-wide scale and revealing potential biological mechanisms [9]. Recently, Parrish et al. developed a new Stacked Regression Transcriptome-Wide Association Study (SR-TWAS) method using data from the GTEx project (Version 8). Through gene expression prediction models across multiple tissues, they identified six AD-related genes in the supplementary motor area and nine Parkinson's Disease (PD)-related genes in the substantia nigra, providing new insights into the genetic mechanisms of these two diseases [10]. Guo et al. applied BGW-TWAS and aggregate Cauchy testing methods to integrate cis- and trans-eQTL data from brain and blood tissues with GWAS summary statistics, identifying 141 TWAS risk genes for AD dementia [11]. Mews et al. utilized the OTTERS TWAS pipeline, combining cis-eQTL data from cortical brain tissue (MetaBrain) and blood (eQTLGen), discovering and validating five new AD-associated genes (PRKAG1, C3orf62, LYSMD4, ZNF439, SLC11A2) as well as six genes near known GWAS loci [12]. TWAS analyses using multi-tissue data provide a novel perspective for understanding the functional impact of AD-associated variants,

further expanding our understanding of the disease's genetic mechanisms. However, while TWAS can reveal associations between gene expression and disease risk, it does not directly establish causality [9].

In large-scale transcriptomic studies, the GTEx project is one of the most comprehensive datasets available, covering gene expression patterns and regulatory mechanisms across nearly 1,000 individuals and multiple tissue types [13]. The UK Biobank (UKB) provides genotype and phenotype data for hundreds of thousands of individuals, offering a rich resource for studying complex traits and diseases [14, 15]. The ROSMAP dataset focuses on AD-related research, containing genotype and gene expression data from brain tissues, serving as an essential foundation for exploring the molecular mechanisms of AD [16].

To further explore the biological mechanisms of diseases, Barbeira et al. developed the PrediXcan method, which combines SNP-based gene expression prediction models with GWAS data to infer gene expression levels and identify disease-related genes [17]. Sun et al., based on the S-PrediXcan method, proposed the UTMOST strategy, integrating multi-tissue gene expression data to identify 53 genes associated with AD risk [8]. PrediXcan not only overcomes the limitations of traditional GWAS but also reveals the mediating role of gene expression in phenotypes. As a widely used method for imputing individual-level transcriptomes from genotype data, PrediXcan estimates gene expression values determined by genetic variants and explores the underlying mechanisms of these associations by testing the relationship between predicted expression values and target phenotypes [18].

In recent years, artificial intelligence (AI), especially deep learning, has been widely applied in biomedical research. By integrating multi-omics data such as gene expression and genotype information, AI captures complex nonlinear relationships, advancing the understanding of disease mechanisms [19]. For example, the GenNet framework developed by Hilten et al. and the DeepGAMI model proposed by Chandrashekar et al. have demonstrated excellent predictive capabilities and the ability to interpret underlying mechanisms in disease research [20, 21]. By integrating gene expression and genotype information through deep learning models, there is potential to enhance the prediction of AD risk, providing new directions for early diagnosis and the discovery of potential therapeutic targets.

Currently, Alzheimer's Disease lacks effective treatment drugs, and traditional drug

development is hindered by high costs, lengthy timelines, and unclear pathogenic mechanisms [22] [23]. Therefore, drug repositioning has emerged as a strategy to accelerate therapeutic development [24]. Many complex trait diseases have risks influenced by the accumulation of variants with low significance levels, posing challenges for drug repositioning [25]. To address this, So et al. proposed a novel strategy that infers gene expression profiles from GWAS data and compares them with drug-induced gene expression changes [26]. This method overcomes the interference of drugs and environmental factors in traditional gene expression studies and has made progress in researching complex diseases such as AD.

In this study, we utilized GWAS summary statistics provided by Jiang et al. [27] and integrated gene expression prediction models from 49 tissues in the GTEx database. Using the S-PrediXcan method, we conducted a comprehensive Transcriptome-Wide Association Study (TWAS) on 61 brain-related traits, ultimately identifying five significant traits, including three related to AD family history. Additionally, we validated the reliability of these significant traits by associating predicted gene expression levels with AD-related phenotypes using gene expression and genotype data from 553 samples of the dorsolateral prefrontal cortex in the ROSMAP database through individual-level PrediXcan analysis. Notably, we proposed an innovative deep multi-layer information fusion model, AD-MIF. This model aims to leverage high-quality multimodal data (such as gene expression and genotype phenotypes) to achieve precise predictions of a series of important phenotypes. Based on deep learning technologies, AD-MIF can comprehensively integrate multimodal data and extract more representative features, significantly enhancing prediction accuracy. Based upon the core regulatory genes of AD identified in our study, we combined the strategy proposed by So et al. to compare GWAS-inferred gene expression profiles with drug-induced gene expression changes, screening a series of potential drug candidates. This strategy provides important clues for exploring AD treatment options and accelerates the development of related drugs. Our study not only offers new perspectives for the early diagnosis and intervention of Alzheimer's Disease but also points out new directions for future research, advancing the further development of this field.

Results

S-PrediXcan Analysis of AD Family History and Related Traits

Gene Expression and Significant Gene Count for AD Family History-Related Traits

We conducted a transcriptome-wide association study (TWAS) using the S-PrediXcan method to analyze gene expression levels across 49 different tissues for 61 traits related to brain and mental health (**Supplementary Table S1**). Data illustrates that, after adjusting for q-values and applying a false discovery rate (FDR) threshold of 0.01, we identified 18 significant traits. We visualized the number of significant genes in different tissues (**Figure 1**). In Figure 1, red and blue represent high and low counts of significant genes, respectively, revealing the pleiotropy of gene expression across various tissues. Notably, traits such as GCST90042665: Father's Disease – Alzheimer's Disease/Dementia, GCST90042678: Mother's Disease – Alzheimer's Disease/Dementia, GCST90042691: Sibling's Disease – Alzheimer's Disease/Dementia, GCST90041866: Irritability, and GCST90041879: Worry/Anxiety exhibit a relatively rich distribution of significant genes.

Among traits related to a family history of Alzheimer's Disease (AD), key brain regions such as the hippocampus (Brain_Hippocampus) and the frontal cortex (Brain_Frontal_Cortex_BA9) demonstrated significant pleiotropy (**Supplementary Figure S2**). This further suggests that these brain regions may play a critical role in the onset and progression of Alzheimer's Disease. Other brain regions, including the anterior cingulate cortex (Brain_Anterior_Cingulate_Cortex_BA24) and the cerebral cortex (Brain_Cortex), showed strong gene associations with emotional traits (e.g., irritability and worry/anxiety), indicating that these areas may be closely related to the emotional and behavioral regulation mechanisms associated with AD. It is noteworthy that the distribution of significant genes is not limited to the brain due to the fact that peripheral tissues such as the liver (Liver), lung (Lung), and arteries (Artery_Aorta, Artery_Coronary) also exhibit a certain number of significant genes. Although their number is lower compared to brain tissues, they exhibit some association with traits linked to a family history of AD (**Figure 2**). This indicates that AD is not solely a neurodegenerative disease; its development may also involve pathological changes in the metabolic and vascular systems, especially in cases with a familial genetic

background.

Overall, there are significant differences in the number of significant genes for different traits within the same tissue. For example, the hippocampus and frontal cortex exhibit a high number of significant genes across multiple AD family history traits and emotional traits, suggesting that these brain regions play a central role in the various genetic mechanisms of AD. In contrast, although peripheral tissues have fewer significant genes, their association with familial AD suggests that these tissues may play important roles in the disease through metabolic or vascular-related pathways.

In summary, the number of significant genes in brain tissues is significantly higher than in non-brain tissues, which aligns with AD being characterized as a neurodegenerative disease. However, the significance observed in peripheral tissues indicates that AD may involve a more extensive systemic involvement, particularly with regard to metabolic, inflammatory, and vascular pathology. Therefore, exploring the synergistic interactions between brain and peripheral tissues could offer new insights into understanding the systemic mechanisms of AD.

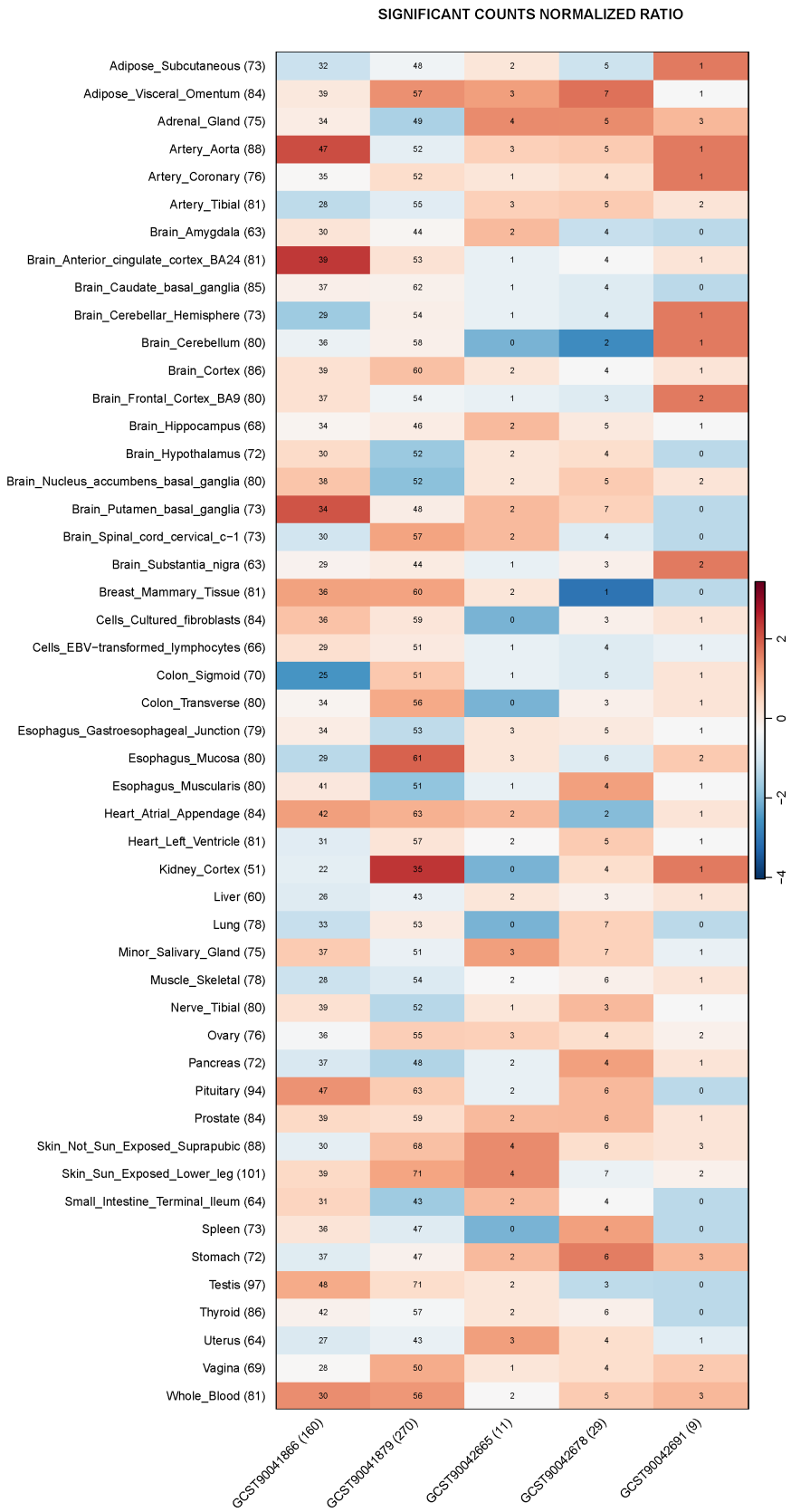


Figure 1. Distribution of Significant Genes for Five Brain and Psychiatric Traits (Including AD Family History and Emotional Traits) Across 49 Tissues Identified via S-PrediXcan Analysis. The heatmap illustrates the distribution of significant genes for five brain- and psychiatric-related traits, selected from an initial analysis of 61 traits, using S-

PrediXcan across 49 different tissues. These traits include: GCST90042665 (Father's Disease: Alzheimer's/Dementia), GCST90042678 (Mother's Disease: Alzheimer's/Dementia), GCST90042691 (Sibling's Disease: Alzheimer's/Dementia), GCST90041866 (Irritability), and GCST90041879 (Worry/Anxiety). Rows represent different tissues, while columns represent these five traits. The colors indicate standardized significant gene counts (Z-scores), with red representing higher gene counts and blue representing lower counts. The results highlight the pleiotropy of these traits and their tissue-specific associations, particularly in brain regions such as the hippocampus and frontal cortex, where significant gene concentrations are notably higher.

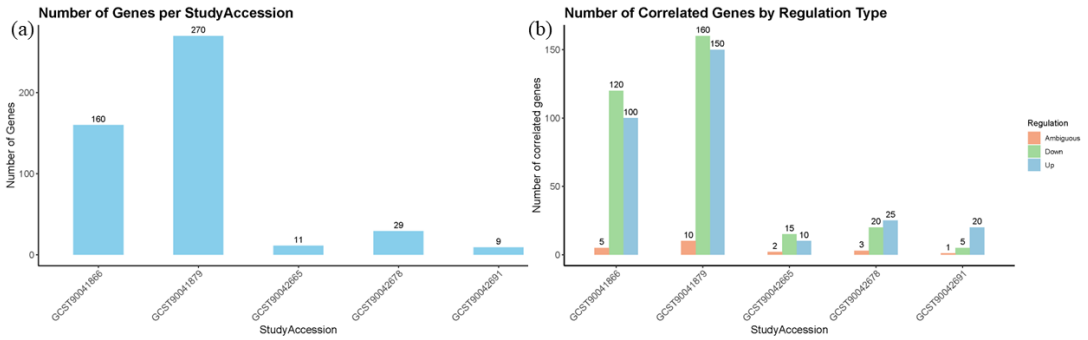


Figure 2. The total number of significant genes and the distribution of up-regulation and down-regulation of gene expression for brain- and psychiatric-related traits. a) Number of significant genes in each trait. This figure displays the total number of significant genes linked to each Alzheimer's Disease (AD)-related trait. The significant genes for each trait are determined based on gene expression differences across multiple tissues. Here, the count represents the number of unique significant genes per trait, meaning each gene is counted only once. b) The Number of Genes Whose Gene Expression is Up-regulated, Down-regulated and Ambiguous in Each Trait. This figure illustrates the number of genes with up-regulated, down-regulated, and ambiguous (i.e., equal numbers of up-regulated and down-regulated) gene expressions for each AD-related trait. The statistics are based on the performance of each gene across multiple tissues. In some cases, a gene may exhibit different expression statuses (up-regulated or down-regulated) in different tissues, resulting in a single gene being counted in multiple categories (up-regulated or down-regulated).

Gene Expression and Enrichment Analysis for AD Family History-Related Traits

Present study conducted a comprehensive analysis of gene expression for five related traits, including a family history of Alzheimer's Disease (AD), using the S-PrediXcan method. Three screening analyses were performed with different q-value thresholds (q-value ≤ 0.05, q-value ≤ 0.1, and q-value ≤ 0.5), and gene network diagrams were constructed. As the q-value threshold was lowered, the selected genes became increasingly concentrated among a few core genes strongly associated with AD. At a q-value ≤ 0.5, the gene distribution was relatively broad, encompassing multiple traits; however, at q-value ≤ 0.1 and q-value ≤ 0.05, the results progressively focused on key genes with strong associations. Detailed information is provided in Supplementary Tables S2–S4.

Gene Expression Networks at q-value ≤ 0.5

In the gene expression network analysis with a q-value ≤ 0.5, we identified shared genes among multiple traits, which are not only closely associated with familial Alzheimer's Disease (AD) but also exhibit significant associations with emotional and behavioral characteristics such as irritability and worry/anxiety. These shared genes reveal complex genetic links between AD family history traits and emotional and behavioral features, indicating their important roles in polygenic regulatory processes. Supplementary Figure S3 illustrates the distribution of significantly associated genes across multiple traits. APOE is a core gene closely related to AD risk, showing strong associations with three family history traits (father's, mother's, and siblings' history of AD/dementia), emphasizing its importance in the genetic mechanisms of AD. APOC1 and SCIMP are not only associated with two familial AD history traits but also significantly related to the trait of irritability, suggesting they may play roles in AD progression and emotion regulation. KAT8 is significantly associated with AD family history, irritability, and anxiety traits, indicating it may have dual functions in emotion regulation and AD genetic mechanisms. Other genes, such as RP11-196G11.6, VKORC1, RP11-196G11.2, PRSS53, and ZNF646, also show associations with familial AD history, irritability, and anxiety traits, further suggesting their potential importance in hereditary AD and emotion regulation. Detailed networks are presented in Supplementary Figure S4.

Gene Expression Networks at q-value ≤ 0.1

When the q-value threshold is set to ≤ 0.1, the number of significantly identified genes decreases, but the associations of these genes become strengthen. In this analysis, we identified core genes strongly associated with multiple familial AD history traits, such as APOE, APOC1, and TOMM40. The gene network becomes more concentrated under this threshold, particularly for traits such as irritability and worry/anxiety, with a notable increase in the number of shared genes. Supplementary Figure S5 shows that these core genes are widely present across multiple familial AD history traits, particularly those related to AD-associated emotional and cognitive symptoms. TOMM40 and APOE show strong associations across all three family history traits, further confirming their central roles in the genetic mechanisms of AD. Other genes, such as RP11-196G11.6, ZNF646, and ZNF668, exhibit shared associations between familial AD history and emotional traits (e.g., irritability, anxiety), suggesting that these genes may play roles in the genetic mechanisms through which emotional and behavioral traits influence AD.

Gene Expression Networks at q-value ≤ 0.05

At a q-value threshold of ≤ 0.05, the number of genes identified is significantly reduced, but almost all the selected genes are closely related to AD, becoming key candidate genes (Figure 3). For example, APOE and NECTIN2 show the strongest associations under this threshold, especially exhibiting higher significance in traits related to familial Alzheimer's Disease/dementia history. These genes are widely involved in the core genetic mechanisms of familial AD. Figure 6 illustrates that the most strongly associated genes are concentrated in a few key traits, particularly showing pronounced significance in traits such as the father's disease – Alzheimer's Disease/dementia and the mother's disease – Alzheimer's Disease/dementia, revealing core genetic loci associated with AD family history. APOE and APOC1 continue to show strong associations with multiple familial AD history traits, while NECTIN2 also exhibits high significance across multiple AD-related traits, further indicating its crucial role in familial AD. Detailed information is available in Supplementary Table S5.

GO and KEGG Enrichment Analysis of AD-Related Genes in Significant Tissues

Based on p-value and q-value ≤ 0.05, we performed GO and KEGG enrichment analyses on genes associated with familial Alzheimer's Disease (AD) traits in tissues such as the aorta (Artery_Aorta), anterior cingulate cortex BA24 (Brain_Anterior_cingulate_cortex_BA24), putamen/basal ganglia (Brain_Putamen_basal_ganglia), esophageal mucosa (Esophagus_Mucosa), and

kidney cortex (Kidney_Cortex). The GO enrichment results (see Figure 4a) indicate that the key genes in these tissues are primarily involved in critical processes such as natural killer cell-mediated cytotoxicity, amyloid β-protein metabolism, and cholesterol transport. These biological processes are closely related to immune responses and lipid metabolism abnormalities in AD. KEGG pathway enrichment analysis (see Figure 4c) further supports these findings, showing significant enrichment of cholesterol metabolism and cell adhesion molecule pathways in primary tissues. This suggests that these tissues play important roles in the genetic and pathological mechanisms of AD.

GO and KEGG Enrichment Analysis of AD-Related Genes in Other Significant Tissues

For tissues outside the primary focus, we also conducted GO and KEGG enrichment analyses based on p-value and q-value ≤ 0.05. The GO enrichment results (see Figure 4b) reveal that the genes in these tissues are mainly involved in processes such as phospholipid efflux, protein-lipid complex organization, and cholesterol transport. These functions are associated with auxiliary metabolic regulation in AD. KEGG pathway analysis (see Figure 4d) indicates that cholesterol metabolism remains the most significantly enriched pathway in non-primary tissues. Although these tissues receive less research attention, their lipid metabolic activity may still hold potential significance in the pathological mechanisms of AD.

Using the screening criteria of p-value and q-value ≤ 0.05, we uncovered functional differences of AD-related genes across various tissues. The key significant tissues play central roles in immune responses and lipid metabolism, while other tissues exhibit roles in auxiliary lipid metabolic regulation. These findings contribute to a deeper understanding of the complex genetic background and multiple pathological mechanisms of AD.

GO and KEGG Enrichment Analysis of AD-Related Genes in the Anterior Cingulate Cortex BA24

To further explore the potential functions of genes associated with familial Alzheimer's Disease (AD) traits in the anterior cingulate cortex (Brain Anterior Cingulate Cortex, BA24), we conducted GO and KEGG enrichment analyses on the selected key genes. The anterior cingulate cortex plays a crucial role in cognitive function, emotion

regulation, and the pathogenesis of AD. Using a screening criterion of p-value ≤ 0.01, we identified multiple biological processes and pathways related to AD pathology.

The GO analysis results (see Figure 5a) show that the relevant genes are mainly enriched in processes such as pigment organization, cell adhesion, detoxification, and natural killer cell-mediated immune responses. Specifically, the APOE and ABCB6 genes are closely related to the formation and metabolism of pigment granules, while the CEACAM19 and PVR genes play key roles in intercellular adhesion and immune functions. Additionally, the enrichment of the APOE gene in processes related to cellular responses to toxic substances further supports its multifaceted role in the pathogenesis of AD. KEGG pathway analysis (see Figure 5c) highlights the central roles of cholesterol metabolism and ABC transporter proteins in AD. The significant enrichment of the APOE gene in the cholesterol metabolism pathway indicates its critical role in regulating lipid metabolism and amyloid protein deposition. The ABCB6 gene is enriched in the ABC transporter protein pathway, potentially affecting AD pathology by regulating the transport of substances in and out of cells. These results suggest that AD-related genes in the anterior cingulate cortex play important roles in lipid metabolism, immune responses, cell adhesion, and detoxification processes, further elucidating their key roles in the complex pathological mechanisms of AD.

GO and KEGG Enrichment Analysis of AD-Related Genes in the Aorta

When investigating the potential functions of genes associated with familial Alzheimer's Disease (AD) beyond brain tissues, we selected the aorta (Artery Aorta) for GO and KEGG enrichment analyses to explore the roles of these genes in peripheral tissues. The analysis revealed several important biological processes and signaling pathways related to cell adhesion, immune responses, and mitochondrial function, using a screening criterion of p-value ≤ 0.01. The findings indicate that these processes are significantly enriched in the aorta, suggesting that the peripheral vascular system may have a significant impact on the pathological mechanisms of AD.

The GO analysis (see Figure 5b) shows that genes such as NECTIN2 and CEACAM19 are closely related to cell–cell adhesion, potentially affecting AD pathology by maintaining vascular structure and function. Additionally, NECTIN2 is enriched in biological processes related to natural killer cell-mediated immune responses, indicating its critical role in immune dysfunction associated with AD. The TOMM40 gene is involved in mitochondrial outer membrane organization and the insertion of

proteins into the mitochondrial outer membrane, suggesting that mitochondrial function may play an important role in the energy metabolism dysregulation in AD. KEGG analysis (see Figure 5d) further supports these findings, showing that the NECTIN2 gene is enriched in adhesion junctions and cell adhesion molecule pathways, emphasizing its crucial role in maintaining vascular wall stability and regulating cell–cell interactions. Additionally, the mitochondrial autophagy pathway of the TOMM40 gene is enriched, implying its significant role in aortic cell homeostasis and AD-related metabolic abnormalities. Overall, these results indicate that AD-related genes in the aorta may have a substantial impact on the pathological progression of AD by regulating cell adhesion, immune responses, and mitochondrial function.

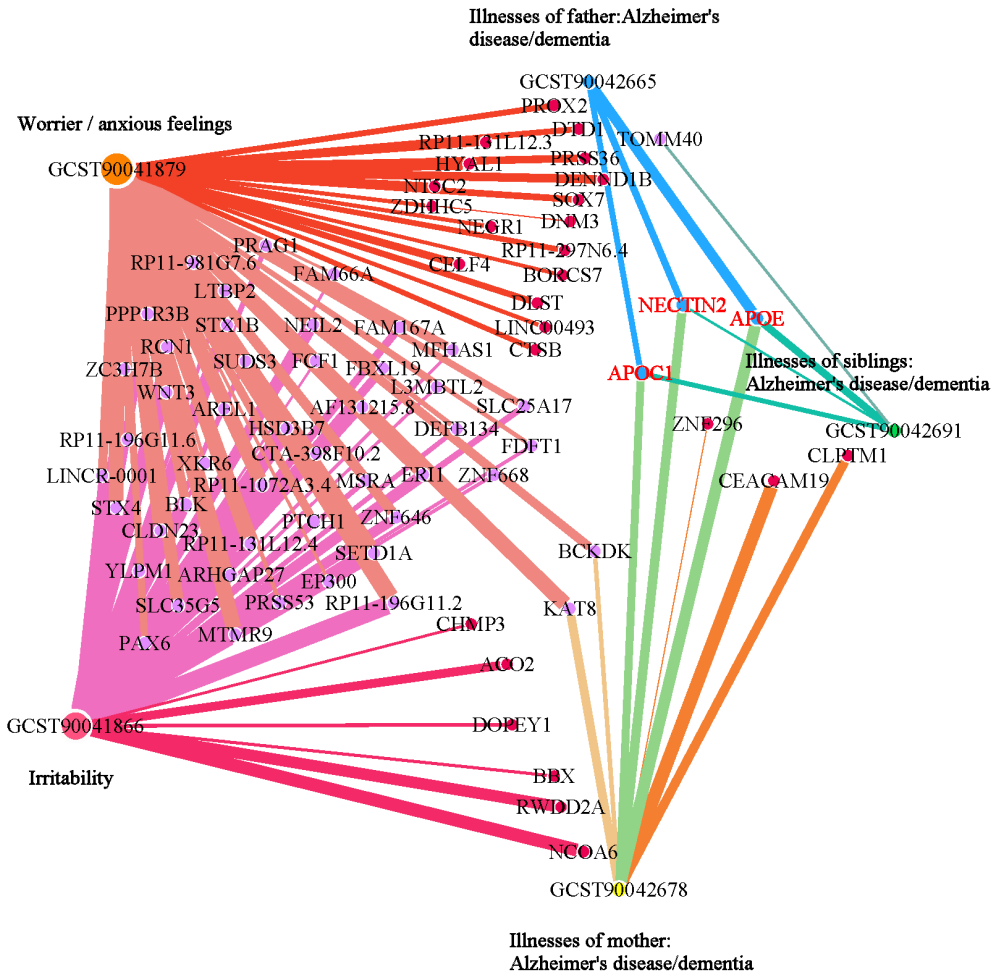


Figure 3. Gene Expression Network for Brain- and Psychiatric-Related Traits at a q-value Threshold of 0.05. This figure illustrates the core genes with the strongest associations identified at a q-value (pva.qval) threshold of 0.05. These genes demonstrate the highest significance in traits related to familial Alzheimer's Disease/dementia history, highlighting their critical roles in the genetic mechanisms underlying AD family history.

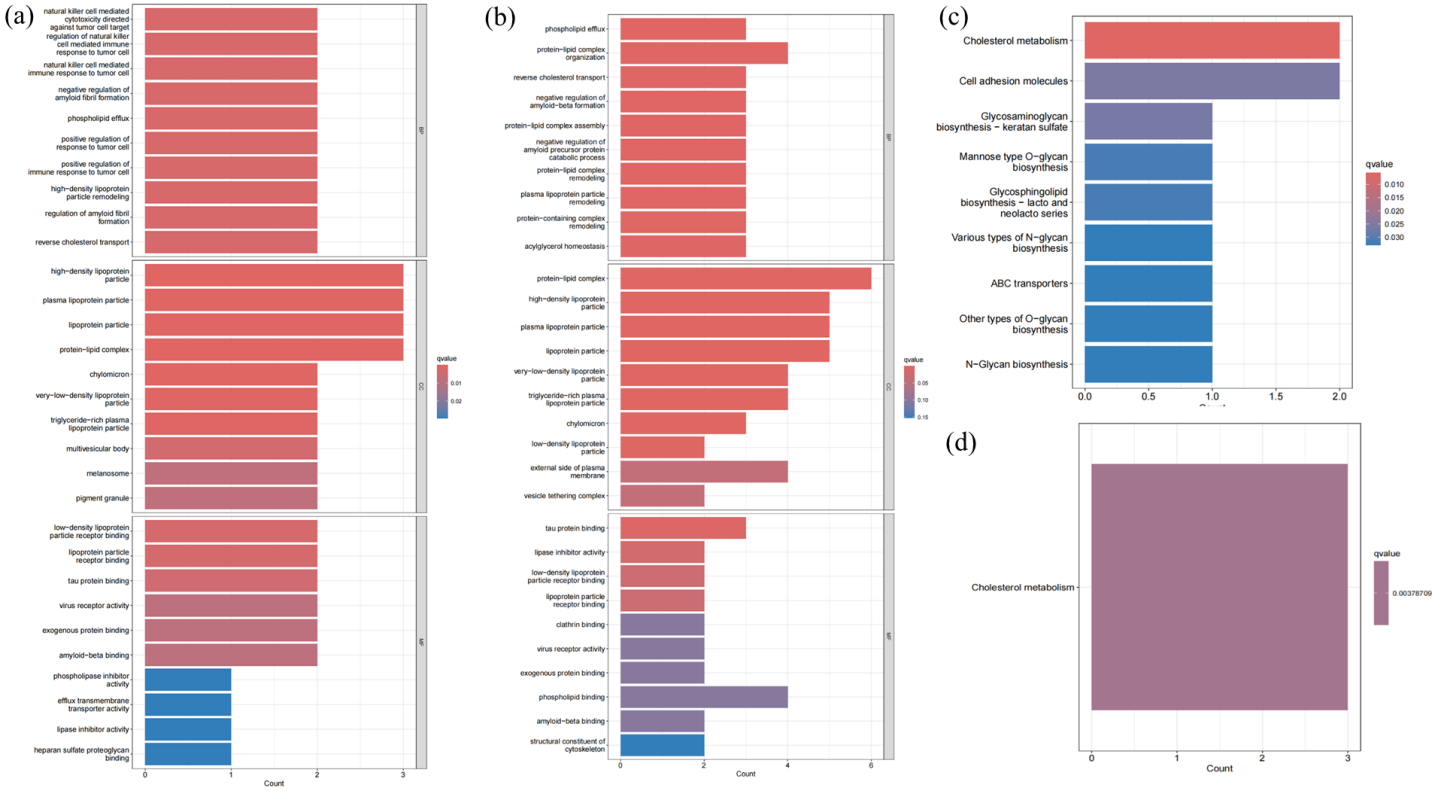


Figure 4. GO Biological Process and KEGG Pathway Enrichment Analysis Across Significant Tissues (p-value and q-value ≤ 0.05). **a)** GO biological process enrichment analysis in key tissues of interest. **b)** GO biological process enrichment analysis in other significant tissues. **c)** KEGG pathway enrichment analysis in key tissues of interest. **d)** GO biological process enrichment analysis in other significant tissues.

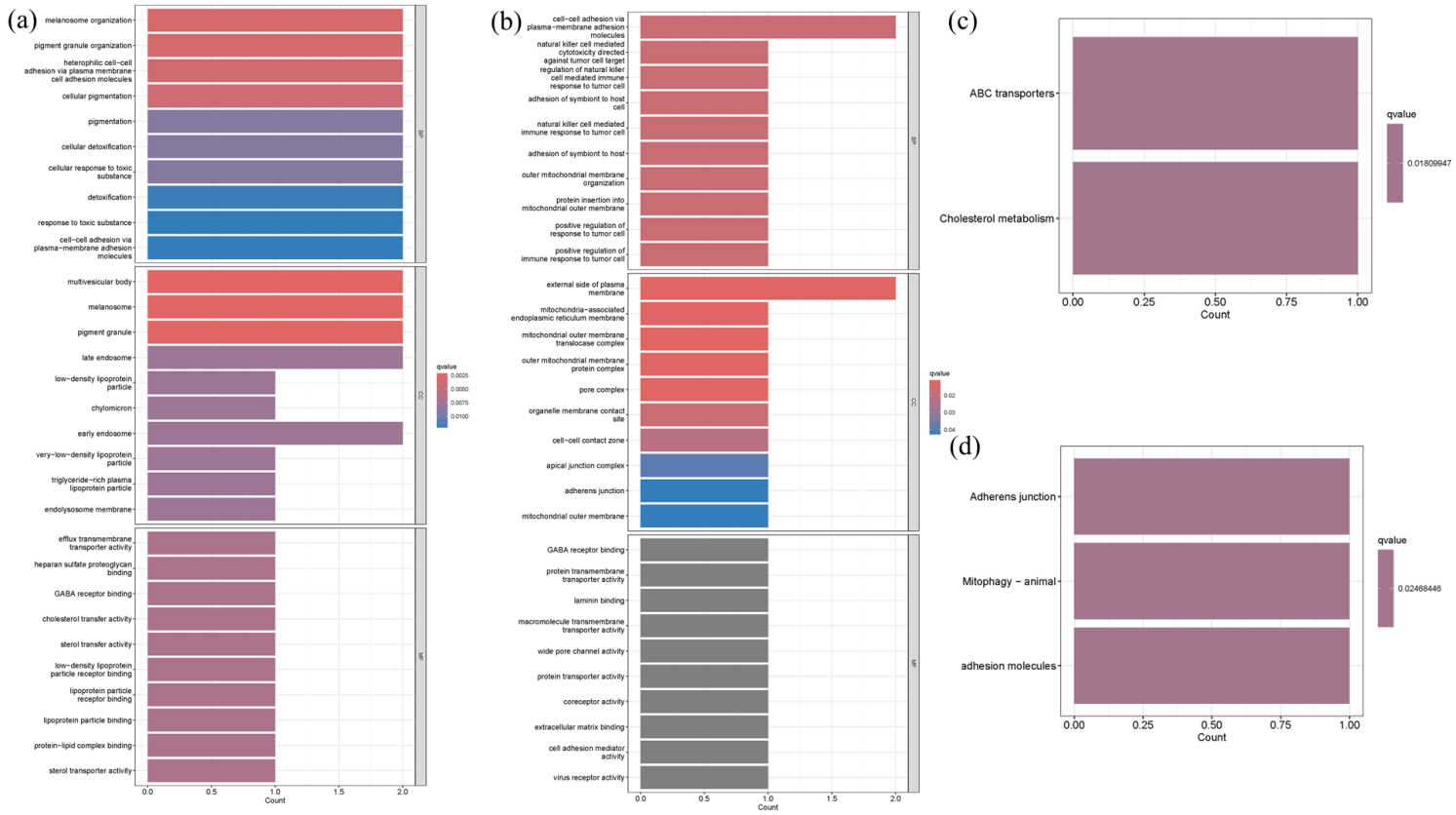


Figure 5. GO Biological Process and KEGG Pathway Enrichment Analysis in Brain Anterior Cingulate Cortex (BA24) and Artery Aorta (q-value ≤ 0.01). **a)** GO Biological Process Enrichment Analysis in Brain Anterior Cingulate Cortex (BA24). **b)** GO Biological Process Enrichment Analysis in Artery Aorta. **c)** KEGG Pathway Enrichment Analysis in Brain Anterior Cingulate Cortex (BA24). **d)** KEGG Pathway Enrichment Analysis in Artery Aorta.

Individual-level PrediXcan Analysis of the ROSMAP Dataset and APOE Association Results

Gene Expression Network Analysis of AD-Related Phenotypes

We conducted a shared gene expression network analysis for nine phenotypes related to Alzheimer's Disease (AD) (phenotype annotation information is provided in Supplementary Table S6). Among these, the APOE genotype is closely associated with the genetic risk of AD and, as a key genetic marker, its impact on AD pathology has garnered significant attention. As illustrated in Supplementary Figure S6, the shared network analysis of gene expression across phenotypes revealed several key genes shared among multiple phenotypes. Notably, the genes NDUFA13, C17orf58, and BTG2 are shared across multiple AD phenotypes, including age at death, MMSE scores at the last follow-up, and Braak pathological staging. The presence of these genes suggests that they may play multiple roles at different stages of AD progression (**Supplementary Table S7**).

For instance, NDUFA13 plays a crucial role in mitochondrial function and energy metabolism, which may be closely related to the metabolic dysregulation observed in AD. C17orf58 and BTG2 are involved in the regulation of cell proliferation and apoptosis, indicating that they may play critical roles in the survival and death processes of neural cells. The shared nature of these genes highlights their central roles in regulating multiple AD phenotypes and provides clues for further functional validation.

Functional Enrichment Analysis of Phenotype-related Genes in Cognitive Diagnosis

To further investigate cognitive diagnostic scores closely associated with the risk of Alzheimer's Disease (AD), we conducted enrichment analyses on the related genes. This revealed the significant functions of multiple genes across different cellular locations, tissues, and subcellular structures, particularly highlighting their potential roles in cognitive function and the nervous system.

Through Gene Ontology (GO) analysis (**Supplementary Figure S7**), several GO terms, such as "intracellular anatomical structure," "membrane-bound organelle," and "cytoplasm," exhibited a significant enrichment of genes (e.g., MTMR7, PPP1R13B, UQCRC1) that play critical roles in fundamental biological processes like neural signal

transduction, metabolism, and protein synthesis. These genes may participate in the cognitive regulation of the nervous system by modulating cellular structures and functions. Additionally, Tissue Expression analysis revealed that multiple genes are highly expressed in key tissues, including the brain, nervous system, and endocrine glands, further suggesting their close association with neural development, neuroprotection, and cognitive function. Subcellular Localization analysis further demonstrated that these enriched genes are widely distributed across various intracellular regions, particularly in the cytoplasm and organelles, and are involved in key processes such as metabolism, protein synthesis, and cell signal transduction. These processes are crucial for neuronal function, synaptic plasticity, and cognitive health. UniProt Keywords analysis revealed post-translational modifications of these genes, including splice variants, phosphorylation, and acetylation, which may play important roles in synaptic plasticity and neural signal regulation by modulating protein stability and activity (Supplementary Table S8).

In summary, individual-level PrediXcan analysis combined with enrichment results identified key genes shared across multiple phenotypes, particularly emphasizing the core roles of NDUFA13, C17orf58, and BTG2. Furthermore, the significant roles of the enriched genes in various cellular functions and processes suggest that they may influence cognitive functions of the nervous system through multiple mechanisms. These genes are not only potential biomarkers but also may serve as therapeutic targets for Alzheimer's Disease.

Performance Evaluation of Comparison Model AD-MIF

In this study, the multimodal AD gene data were randomly divided into training sets (80%), validation sets (10%), and test sets (10%). These subsets were used for the training and evaluation of the AD-MIF model. To assess the predictive performance of AD-MIF on various downstream tasks for AD patients, such as cognitive level, Braak Stage, and APOE, we compared AD-MIF with general machine learning and deep learning prediction models based on omics data. To ensure the reliability of the results, the baseline models employed the same data splitting method as AD-MIF and used the average values from five-fold cross-validation as the final outcomes. Regarding performance evaluation metrics, we focused primarily on the Area Under the Curve (AUC) because it provides a comprehensive assessment of the model’s classification performance under different conditions.

As shown in Table 1, AD-MIF achieved the best average performance in predicting various downstream tasks related to AD. Compared to traditional machine learning algorithms such as Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF), AD-MIF exhibited an average increase in AUC of 10%-20% across different downstream tasks. This indicates that AD-MIF can effectively capture and utilize more complex feature information, demonstrating a stronger ability to differentiate gene expression and genotype features associated with Alzheimer's Disease. Compared to deep learning algorithms like ResNet and Recurrent Neural Networks (RNN), AD-MIF achieved an average AUC improvement of over 15%, suggesting that AD-MIF's specific fusion modules, which integrate potential feature information from multiple omics data, further optimize feature representation and model learning. Notably, when compared to modality imputation deep learning methods such as DeepGAMI and GenNet, AD-MIF also improved accuracy by approximately 1%-5%. This suggests that AD-MIF enhances the model's generalization ability and accuracy by establishing more effective associations between original omics features and neighborhood features.

In summary, AD-MIF enhances the expression capability and information utilization efficiency of different modal features by integrating and learning various omics characteristics. This provides robust theoretical and technical support for the practical applications of AD-MIF in early detection, progression prediction, and personalized treatment of Alzheimer's Disease.

Additionally, to further examine the effectiveness of various configurations of AD-MIF, this study established the following AD-MIF variants:

- AD-MIF without genotype data (w/o GT): Removed the genotype data input;
- AD-MIF without genotype data (w/o GT): Removed the genotype data input;
- AD-MIF without modal fusion (w/o MF): Removed the modal fusion algorithm.

All AD-MIF variants were processed using the same data splitting method and their effectiveness was evaluated using the AUC metric. The experimental results are presented in Table 2. The AD-MIF variant without genotype data (w/o GT) resulted in a decrease in model accuracy of approximately 5.5% across all AD-related prediction tasks. This indicates that AD-related gene phenotype data provide richer biological guidance information to the model, enabling it to more comprehensively capture the complex biological features of AD and improve prediction accuracy. The results of

removing the genetic/SNP neighborhood feature graph representation (w/o Gen-G/SNP-G) demonstrate that various interaction relationships between neighborhoods can enhance AD-MIF's ability to represent and utilize features extracted from omics data efficiently. Notably, the AD-MIF variant without modal fusion (w/o MF) caused the largest decrease in model prediction accuracy, with an average reduction of 7.3%. This suggests that the information fusion module employed by AD-MIF, by integrating local and global features, generates more consistent feature representations, providing higher-quality features for AD-related tasks and enhancing the model's prediction accuracy and stability.

In summary, AD-MIF enhances the representation capability and information utilization efficiency of different modal features by integrating and learning various omics data characteristics. This provides robust theoretical and technical support for the practical applications of AD-MIF in early detection, progression prediction, and personalized treatment of Alzheimer's Disease.

Table 1. Performance of AD-MIF

Methods	Cognitive Level	Cerad Score	Braak Stage	APOE
SVM	0.544	0.477	0.532	0.502
LR	0.575	0.475	0.498	0.486
RF	0.566	0.465	0.572	0.472
ResNet	0.601	0.496	0.502	0.455
RNN	0.636	0.499	0.501	0.504
ML-KNN	0.538	0.475	0.542	0.559
DeepGAMI	0.621	0.550	0.589	0.642
GenNet	0.635	0.542	0.591	0.677
AD-MIF	**0.642±0.015**	**0.558±0.023**	**0.602±0.011**	**0.721±0.024**

Table 2. AD-MIF Ablation Experiment Results

Methods	Cognitive Level	Cerad Score	Braak Stage	Apoe
AD-MIF(w/o GT)	0.582	0.552	0.534	0.634
AD-MIF(w/o Gen-G /SNP-G)	0.560	0.557	0.531	0.584
AD-MIF(w/o MF)	0.561	0.544	0.530	0.592
AD-MIF	**0.642±0.015**	**0.558±0.023**	**0.602±0.011**	**0.721±0.024**

Visualization and Analysis of AD-MIF Model Results Based on Gene Expression Data

To investigate the specific roles that related genes play in AD-related phenotype prediction tasks, the experiment extracted neighborhood feature graph representations of gene expression under different AD tasks (Gen-G) and analyzed the importance scores of each gene in various tasks. Specifically, the AD-MIF model calculated the attention weights for each gene based on Gen-G. By analyzing these weight scores, we can gain deeper insights into the relative contributions of genes across different tasks, thereby revealing their biological significance and potential functional roles in AD prediction. As shown in Figure 6, the results indicate that different genes make significant contributions to various phenotypes of Alzheimer's Disease (AD) (cognitive level, CERAD score, Braak stage, and APOE genotype). Key genes such as ADRA1A, CAMKK1, SLC30A4, and PNKD are likely closely related to neural signal transmission, amyloid plaque formation, neurofibrillary tangle progression, and APOE-related risk regulation. These genes provide essential foundations for understanding the phenotypic differences of AD, its underlying mechanisms, and the exploration of diagnostic and therapeutic targets.

Additionally, to explore the feasibility and effectiveness of artificial intelligence methods in studying disease mechanisms, we conducted enrichment analyses on key genes closely related to cognitive diagnostic scores associated with Alzheimer's Disease (AD) risk (Supplementary Table S9). The genes predicted by the model to be associated with cognition were significantly enriched in Gene Ontology (GO) components such as intracellular anatomical structures (e.g., organelle maintenance and signal transduction), nucleoplasm, and nuclear lumen. This indicates that these genes play crucial roles in key biological processes such as gene expression regulation, signal transduction, and cellular function maintenance, thereby supporting normal neuronal functions and cognitive processes (Supplementary Figure S8). Furthermore, tissue enrichment analysis revealed that these genes are highly expressed in the brain and nervous system, further confirming their critical roles in neurosignal transmission, synaptic plasticity, and neuronal development—biological processes related to cognition. These findings are consistent with previous studies, emphasizing the importance of intracellular, particularly nuclear, functions and gene expression in the

central nervous system for cognitive regulation. This provides a solid foundation for the in-depth exploration of the molecular mechanisms underlying cognitive processes.

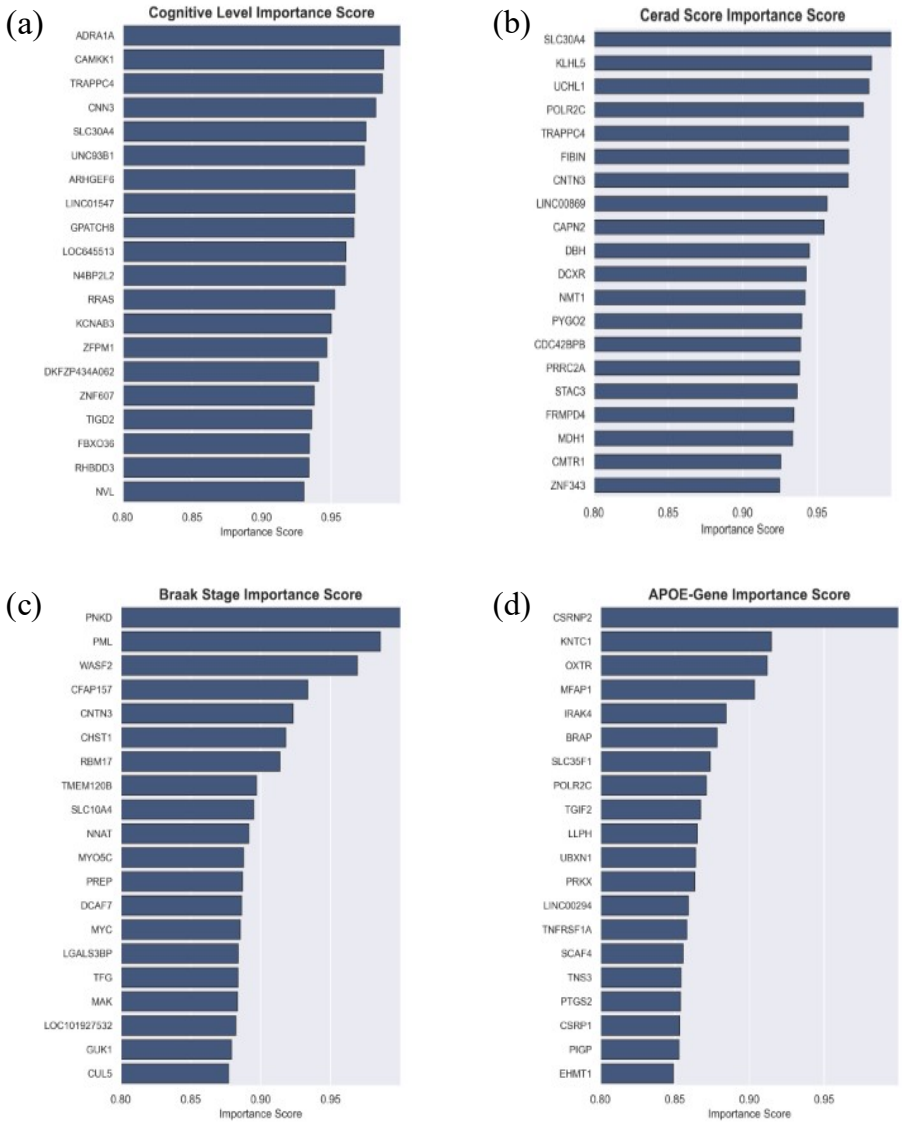


Figure 6. Importance Scores of Genes Across Different AD Phenotype-Related Prediction Tasks. **a)** Importance scores of genes in the Cognitive Level prediction task; **b)** Importance scores of genes in the CERAD Score prediction task; **c)** Importance scores of genes in the Braak Stage prediction task; **d)** Importance scores of genes in the APOE prediction task.

Drug Repositioning Analysis and Potential Therapeutic Drugs for Alzheimer's Disease

Drug Repositioning Analysis for AD Family History-Related Traits

This study utilized the S-PrediXcan method to analyze the gene expression levels associated with three Alzheimer's Disease (AD) family history traits (history of AD/dementia in father, mother, and siblings). We identified significant genes with $p \leq 0.01$ and conducted drug repositioning analysis. By calculating q-values and performing manual filtering, we extracted the top 50 most promising drugs from each trait and tissue. The results are presented in heatmaps and bar charts. Additionally, the top 20 most significant candidate drugs for each AD family history trait are listed in Table 3, Supplementary Table S10, and Supplementary Table S11. Intersection analysis further revealed common drugs and tissues across the three AD family history traits, indicating that these drugs may have significant potential in treating multiple AD-related symptoms (Supplementary Figure S9).

Drug Repositioning Analysis for Maternal Alzheimer's Disease/Dementia History

In the drug repositioning analysis for maternal Alzheimer's Disease (AD)/dementia family history, drugs such as Dasatinib and Sirolimus exhibited significant potential in tissues like whole blood and the aorta (see Figure 7). Dasatinib shows promise in AD treatment by clearing senescent cells, thereby alleviating neurodegeneration and associated mood symptoms. Sirolimus, on the other hand, slows neuroinflammation in AD by modulating immune responses, further validating its broad potential in various neurodegenerative diseases. Additionally, drugs like Guanidine, Mianpril, and others also demonstrated the possibility of treating AD in these tissues, particularly showing positive effects in regulating the vascular and circulatory systems.

Drug Repositioning Analysis for Paternal Alzheimer's Disease/Dementia History

In the drug repositioning analysis for paternal Alzheimer's Disease (AD)/dementia history, drugs identified in adrenal gland tissue such as Transitro, Isoxsuprine, and Tamoxifen also showed potential roles in treating AD or cognitive decline (Supplementary Figure S10). The mechanisms of these drugs may be related to improving metabolic functions and regulating blood pressure, which could have positive impacts on AD-related vascular lesions and the cognitive impairments they

cause.

Drug Repositioning Analysis for Sibling Alzheimer's Disease/Dementia History

In the drug repositioning analysis for sibling Alzheimer's Disease (AD)/dementia history (Supplementary Figure S11), drugs found in gastric tissue, including Catechin, Imipramine, and Pioglitazone, demonstrated significant repositioning potential in the digestive system and metabolic regulation. Imipramine, a commonly used antidepressant, plays a positive role in AD's pathological processes by reducing inflammatory responses and improving neuroprotection. The identification of these drugs provides potential therapeutic targets, especially in alleviating AD symptoms by regulating metabolism and reducing neuroinflammation.

Furthermore, based on drug repositioning analyses for emotion-related features such as irritability and anxiety/worry, several drugs showed significant repositioning potential across multiple tissues, including Pantoprazole, Mitoxantrone, and Balaglitazone (see Supplementary Figures S12 and S13). Further analysis results are presented in Supplementary Tables S12 and S13, which list the drug repositioning outcomes for irritability and anxiety/worry based on Q-values and P-values rankings, respectively. These drugs exhibit important potential in regulating neurological functions and cardiovascular metabolism, offering new possibilities for managing patients' emotional symptoms.

The drug repositioning analysis also revealed the reutilization value of various existing drugs. For example, Nintedanib, primarily used to treat fibrotic diseases, may exhibit neuroprotective effects in AD by reducing myelin oligodendrocyte cell degeneration. Quizartinib, by inhibiting RIPK1 signaling, reduces inflammatory responses, suggesting its potential application in neuroinflammation. Additionally, MK-2206, an Akt inhibitor, may protect against neurodegenerative changes in AD by regulating cell survival and apoptosis. Drugs like Sirolimus and Dasatinib showed significant efficacy across multiple AD-related phenotypes, particularly excelling in immune regulation and clearing senescent cells. The multiple mechanisms of action of these drugs provide strong support for their application in AD treatment.

The drug repositioning analysis also revealed the reutilization value of various existing drugs. For example, Nintedanib, primarily used to treat fibrotic diseases, may exhibit neuroprotective effects in AD by reducing myelin oligodendrocyte cell degeneration.

Quizartinib, by inhibiting RIPK1 signaling, reduces inflammatory responses, suggesting its potential application in neuroinflammation. Additionally, MK-2206, an Akt inhibitor, may protect against neurodegenerative changes in AD by regulating cell survival and apoptosis. Drugs like Sirolimus and Dasatinib showed significant efficacy across multiple AD-related phenotypes, particularly excelling in immune regulation and clearing senescent cells. The multiple mechanisms of action of these drugs provide strong support for their application in AD treatment.

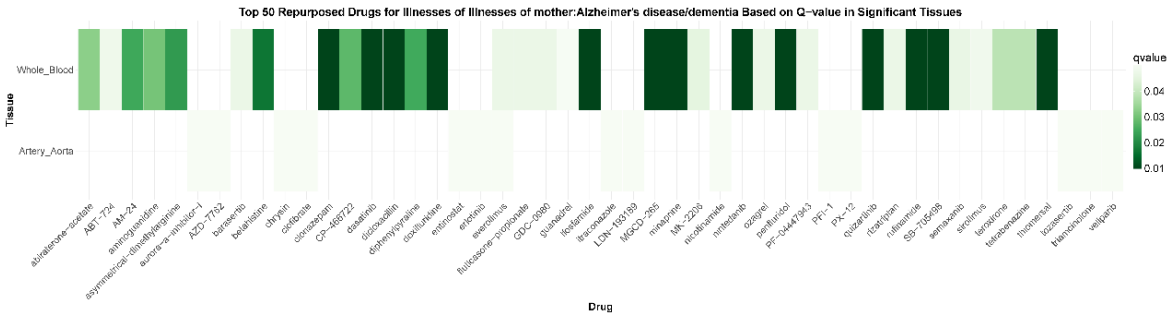


Figure 7. Top 50 Repurposed Drugs for Mother's Alzheimer's Disease/Dementia Based on Q-value in Significant Tissues. This figure illustrates the top 50 potential repurposed drugs identified in significant tissues (Whole Blood and Artery Aorta) for maternal Alzheimer's Disease (AD)/Dementia history, visualizing the drug effects based on q-values. Due to the small sample size and the presence of numerous duplicate values in the data, some correlation coefficients could not be calculated; these data were saved in non-significant tissue files.

Table 3 Top Repositioning Hits for Mother's Alzheimer's Disease/Dementia (Ranked by Q-values and P-values)

Drug	Tissue	Rank	P-value	Q-value	Z-score	Brief description
MGCD-265	Whole Blood	1	9.31e-08	9.92e-03	-1.756	Identified only through virtual screening, no psychiatric relevance.
Nintedanib	Whole Blood	2	1.86e-07	9.92e-03	-1.756	Identified as AD repurposing candidate due to proximity to oligodendrocytes disease modules[28].
Quizartinib	Whole Blood	3	2.79e-07	9.92e-03	-1.756	No psychiatric relevance. Protects against TNF-induced inflammation

						in mice via *Ripk1* inhibition[29].
SB-705498	Whole Blood	4	3.72e-07	9.92e-03	-1.756	No psychiatric relevance, only related to TRP channels.
Dasatinib	Whole Blood	5	4.66e-07	9.92e-03	-1.756	Targets tyrosine kinase activity, key for AD. Clarifies microglial activation. Promising in age-related disease therapies[30].
Clonazepam	Whole Blood	6	5.59e-07	9.92e-03	-1.756	Reduces NO release and microglial activity in LPS-induced conditions[31].
Rufinamide	Whole Blood	7	6.52e-07	9.92e-03	-1.756	Suppresses neuronal hyperexcitability, approved for Lennox-Gastaut Syndrome[32].
Doxifluridine	Whole Blood	8	7.45e-07	9.92e-03	-1.756	No known relevance, identified through virtual screening.
Thiomersal	Whole Blood	9	8.38e-07	9.92e-03	-1.756	No psychiatric relevance. Controversial mercury-based preservative, possible toxicity.
Dicloxacillin	Whole Blood	10	9.31e-07	9.92e-03	-1.756	No psychiatric relevance. Effective against MSSA bacteraemia with a good safety profile[33].
Minaprine	Whole Blood	11	1.02e-06	9.92e-03	-1.756	No psychiatric relevance, identified through virtual screening.
Penfluridol	Whole Blood	12	1.12e-06	9.92e-03	-1.756	An antipsychotic drug with potential anti-cancer properties[34].
MK-2206	Whole Blood	13	1.21e-06	9.92e-03	-1.756	No psychiatric relevance. Inhibits cell viability in DLBCL lines via Akt inhibition[35].

Ifosfamide	Whole Blood	14	1.30e-06	9.92e-03	-1.756	Identified through virtual screening, no further relevance.
Betahistine	Whole Blood	15	2.33e-06	1.65e-02	-1.756	May reduce vertigo symptoms, generally well-tolerated[36].
Asymmetrical-Dimethylarginine	Whole Blood	16	3.35e-06	2.23e-02	-1.756	No psychiatric relevance. Endogenous NO synthesis inhibitor.
AM-24	Whole Blood	17	3.91e-06	2.45e-02	-1.756	No psychiatric relevance.
Diphenylpyraline	Whole Blood	18	4.19e-06	2.48e-02	-1.756	No psychiatric relevance. Inhibits Ebola virus proliferation[37].
CP-466722	Whole Blood	19	4.93e-06	2.77e-02	-1.756	No psychiatric relevance. Identified as an ATM inhibitor, no psychiatric relevance.
Aminoguanidine	Whole Blood	20	5.68e-06	3.02e-02	-1.756	No psychiatric relevance. Potent AGE inhibitor, alleviates diabetic complications[38].

'Rank' refers to the ranking within each tissue for the disorder under study.

Drug Repurposing Analysis through Trait-Tissue-Gene-Drug Networks

To further validate the reliability of the drug repositioning results, this study conducted an in-depth analysis of the potential drugs identified for Alzheimer's Disease (AD) family history-related traits by constructing trait-tissue-gene-drug interaction networks. These networks integrated gene expression data from the LINCS L1000 database, providing systematic insights into AD drug repositioning. By analyzing drugs with a significance level of $q \leq 0.001$ associated with maternal, paternal, and sibling AD-related traits, we revealed the action pathways of these drugs across various AD-related tissues. Interestingly, the reverse targeting results differed from the forward drug repositioning results, indicating drug-tissue specificity and distinct gene-associated pathways.

In the interaction network for maternal Alzheimer's Disease traits, tissues such as the aorta, adrenal glands, uterus, and whole blood exhibited significant gene-tissue associations (Figure 8). Key genes include CLPTM1, NECTIN2, APOE, and TOMM40, which are closely related to the pathological processes of AD. These genes may play important roles in the onset and progression of AD by regulating biological processes such as cholesterol metabolism, immune responses, and cell adhesion. For example, TOMM40 showed strong significance in the aorta, suggesting its potential core role in regulating vascular function and metabolism.

In the network for paternal Alzheimer's Disease traits, adrenal gland tissue showed close interactions with key genes APOE, APOC1, and NECTIN2 (Supplementary Figure S14). APOE and APOC1 play crucial roles in regulating lipid metabolism and clearing amyloid plaques, which are particularly important for the pathogenesis of AD. The associations of these genes with multiple drugs were also significant, suggesting that these drugs may have therapeutic potential in reducing neuroinflammation and regulating metabolism. Through network analysis, we were able to visualize how these drugs interact with key genes in specific tissues, further supporting the potential application of these drugs in AD treatment.

The interaction network for sibling Alzheimer's Disease traits indicated that the stomach and adrenal glands are important target tissues (Supplementary Figure S15). Key genes associated with these tissues include TOMM40, APOE, APOC1, and ABCB6. TOMM40 and APOE exhibited strong significance in their interactions with

drugs across multiple tissues, especially in the stomach. These genes may play dual roles in the pathological processes of AD by regulating metabolism and immune functions. Additionally, the enrichment of the mitochondrial autophagy pathway for the ABCB6 gene suggests its potential role in cellular homeostasis and metabolic regulation. These findings provide new targets for AD treatment, particularly by improving metabolism and enhancing cellular homeostasis to alleviate the disease.

Interaction networks for emotion-related traits such as irritability and anxiety/worry also demonstrated the potential for drug repositioning. Supplementary Figures S16 and S17 detail the trait-tissue-gene-drug interaction networks for these traits. The results indicate that some drugs show consistent significance across multiple tissues, particularly in the nervous and cardiovascular systems. The multi-tissue and multi-gene associations of these drugs suggest that they may influence emotional symptoms through multiple pathways, offering new therapeutic approaches for managing emotions in AD patients.

These trait-tissue-gene-drug networks reveal the complexity of the pathological mechanisms of Alzheimer's Disease, emphasizing the importance of adopting tissue-specific approaches in drug repositioning. The results of the network analysis provide insights into how to utilize gene-tissue-specific interactions for drug screening, thereby facilitating the development of personalized treatment plans. These findings also suggest that future research should place greater emphasis on tissue and gene specificity to more accurately select targeted drugs, thereby enhancing the efficacy and safety of AD treatments.

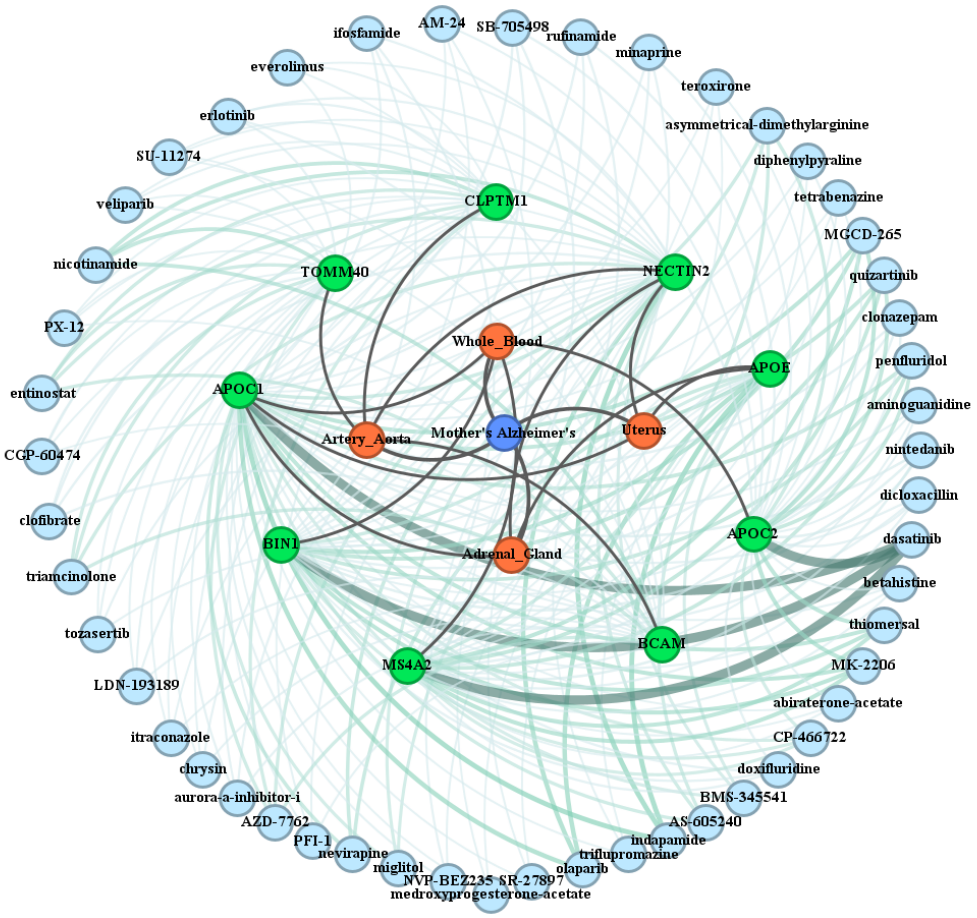


Figure 8. Mother's Alzheimer's Disease: Trait-Tissue-Gene-Drug Interaction Network. This network visualizes the interactions between Mother's Alzheimer's Disease, various tissues, key genes, and potential repurposed drugs. The tissues involved are Artery Aorta, Adrenal Gland, Uterus, and Whole Blood, which play important roles in gene-tissue interactions related to Alzheimer's disease. Key genes include CLPTM1, NECTIN2, APOE, TOMM40, APOC1, BIN1, MS4A2, BCAM, and APOC2, all linked to Alzheimer's pathology. Drug-Gene interactions are weighted by $-log_{10}(p)$ to reflect the strength of the associations, while Tissue-Gene interactions have a fixed weight of 8. This network offers insights into drug repurposing strategies by focusing on crucial tissue and gene interactions in Alzheimer's disease.

Drug Repositioning Analysis for APOE Genotype

In this study, we focused on significant genes associated with the APOE genotype and conducted drug repositioning analyses on these genes (Supplementary Figure S18). By performing a detailed analysis of the genes significantly associated with the APOE genotype, we identified a series of drugs with potential therapeutic effects for Alzheimer's Disease (AD) (Table 4). These drugs may effectively intervene in APOE genotype-related AD pathological processes through various pathways, including the regulation of metabolism, immune responses, and neuroprotection.

In the dorsolateral prefrontal cortex, we identified several important drugs through the gene-drug interaction network (Figure 9). These drugs showed significant associations with key genes such as APOE and TOMM40. Notably, Sirolimus and Dasatinib exhibited strong significance across multiple genes, potentially playing critical roles in alleviating neuroinflammation and clearing senescent cells. Additionally, Everolimus and Temsirolimus demonstrated potential in AD treatment by regulating the mTOR signaling pathway, which helps mitigate neurodegenerative lesions.

During the drug screening for the APOE genotype, we also identified some drugs that may possess specific neuroprotective effects. For example, Gefitinib, an EGFR inhibitor, showed potential in reducing AD pathology and improving cognitive function. Pantoprazole, a proton pump inhibitor, demonstrated a role in decreasing AD risk in follow-up studies. Moreover, preclinical evidence supporting Dasatinib's ability to clear senescent cells further endorses its potential application as a therapeutic agent for AD.

The significance analysis of these drugs was weighted based on $-\log_{10}(p)$ to reflect the strength of the association between the drugs and the genes. By visualizing the multiple action pathways of these drugs in specific tissues, we can better understand their therapeutic mechanisms. These findings highlight the potential of drug repositioning in a personalized context, especially for patients with the APOE genotype, where these drugs may offer more effective intervention methods.

In summary, through drug repositioning analysis related to the APOE genotype, we identified a series of drugs with potential therapeutic effects. These drugs may provide effective treatment options for AD patients by regulating metabolism, reducing neuroinflammation, and clearing senescent cells through various mechanisms. Future research should further explore the in vivo and in vitro experimental validation of these

drugs and evaluate their safety and efficacy in patients with specific APOE genotypes through clinical trials, thereby advancing the development of personalized treatments.

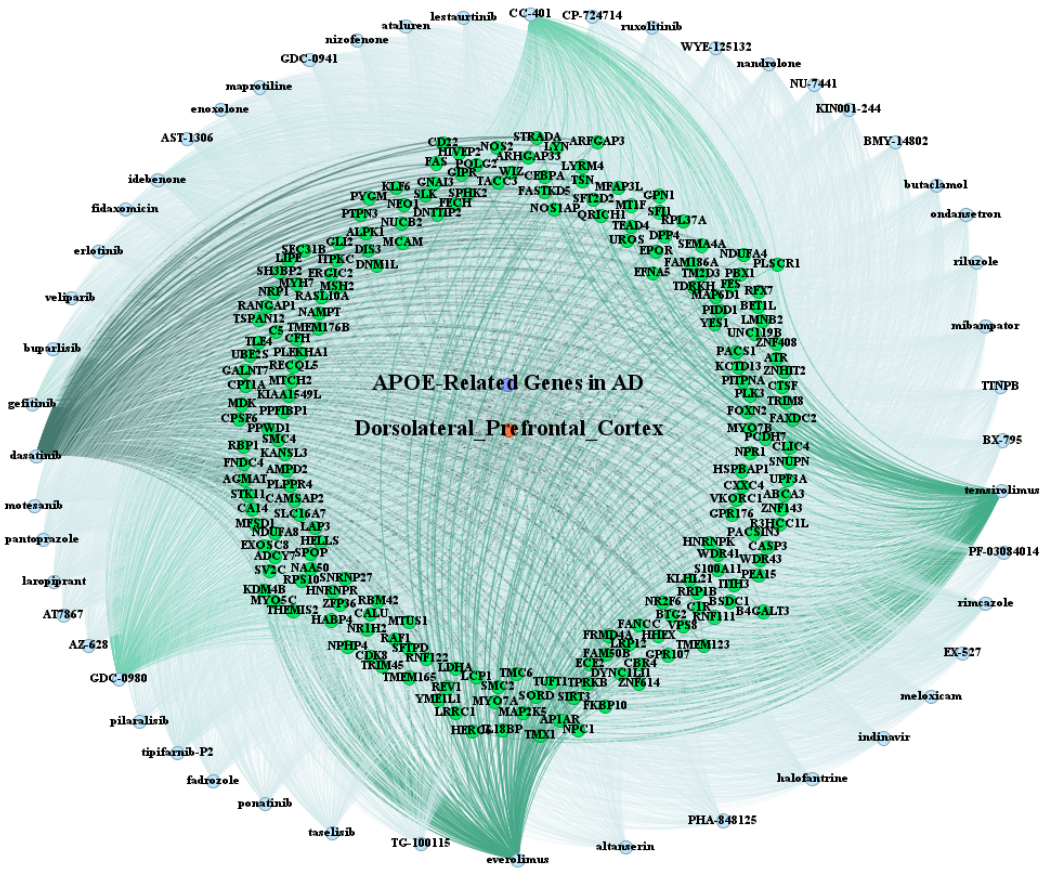


Figure 9. APOE-Related Trait-Tissue-Gene-Drug Network for Alzheimer's Disease Drug Repurposing. This network illustrates interactions between APOE-related traits, tissues, genes, and repurposed drugs for Alzheimer's disease. Nodes represent genes linked to APOE-related phenotypes (from PrediXcan analysis), the dorsolateral prefrontal cortex, and potential repurposed drugs. Edges indicate drug-gene associations weighted by $-log_{10}(p)$ and tissue-gene connections with a fixed weight of 8. Drugs were selected based on significant associations with APOE-related genes ($p \leqslant 0.05$). The network reveals how genetic and tissue-specific interactions inform drug repurposing strategies for Alzheimer's disease.

Table 4 Top Repositioning Hits for APOE Genotype-Related Genes in Alzheimer's Disease/Dementia (Ranked by P-values)

Drug	Tissue	Rank	P-value	Brief description
Gefitinib	Dorsolateral Prefrontal Cortex	1	2.38E-05	EGFR inhibitor, reduces AD pathology and improves cognitive function[39].
Idebenone	Dorsolateral Prefrontal Cortex	1	2.38E-05	Ameliorates memory impairment by promoting Aβ42 clearance and restoring antioxidant status[40].
Rimcazole	Dorsolateral Prefrontal Cortex	2	3.33E-05	No psychiatric relevance.
Tipifarnib-P2	Dorsolateral Prefrontal Cortex	3	4.93E-05	No psychiatric relevance.
BX-795	Dorsolateral Prefrontal Cortex	4	8.25E-05	No psychiatric relevance.
Pantoprazole	Dorsolateral Prefrontal Cortex	5	8.98E-05	Linked to a reduced risk of AD in follow-up studies[41].
Butaclamol	Dorsolateral Prefrontal Cortex	6	1.03E-04	No psychiatric relevance.
Dasatinib	Dorsolateral Prefrontal Cortex	7	1.04E-04	Preclinical evidence for senescent cell clearance, reducing AD pathology[42].
Taselisib	Dorsolateral Prefrontal Cortex	7	1.04E-04	No psychiatric relevance.
Motesanib	Dorsolateral Prefrontal Cortex	8	1.07E-04	No psychiatric relevance.
Everolimus	Dorsolateral Prefrontal Cortex	9	1.16E-04	No psychiatric relevance.
Meloxicam	Dorsolateral Prefrontal Cortex	10	1.34E-04	No psychiatric relevance.
Dasatinib	Dorsolateral Prefrontal Cortex	/	/	/
Temsirolimus	Dorsolateral Prefrontal Cortex	11	1.59E-04	Reduces neurofibrillary tangle density and improves memory[43].
TTNPB	Dorsolateral Prefrontal Cortex	12	1.62E-04	Promotes glia-neuron conversion to replace impaired neurons[44].
TG-100115	Dorsolateral Prefrontal Cortex	13	1.73E-04	No relevant findings.

EX-527	Dorsolateral Prefrontal Cortex	14	1.95E-04	No relevant findings.
CC-401	Dorsolateral Prefrontal Cortex	15	1.98E-04	Trials terminated, possible toxicity.
AST-1306	Dorsolateral Prefrontal Cortex	16	2.01E-04	No relevant findings.
GDC-0980	Dorsolateral Prefrontal Cortex	17	2.26E-04	No relevant findings.

Drug Enrichment Analysis Results

In this study, we conducted a detailed analysis of the enrichment of drug repositioning to evaluate the potential roles of different drug categories in the treatment of Alzheimer's Disease (AD). Utilizing drug sets based on the ATC classification system and MEDI-HPS definitions, we performed various enrichment tests to identify the significance of specific drug categories in AD-related traits.

Drug enrichment analysis based on ATC classification system

The results of the drug enrichment analysis based on the ATC classification system indicated that antiepileptic drugs (N03), antiparkinsonian drugs (N04), antipsychotic drugs (N05A), and anxiolytic/antidepressant drugs (N05B or N06A) exhibited significant enrichment across various AD-related traits. Notably, the significance of these drugs was particularly prominent in emotion-related traits such as irritability and anxiety/worry. These findings suggest that these psychotropic medications may have substantial potential in alleviating AD-related emotional symptoms by regulating neurotransmitters or reducing neuroinflammation.

Table 5 Drug enrichment analyses P values with drug-sets defined in the ATC classification system

Disorder	Self (Fisher)	Compet (Fisher)	Self (minP)	Compet (minP)
Enrichment for ATC antiepileptics（N03）				
Irritability	0	0.653	1.78E-91	0.397
Worrier / anxious feelings	0	1.56E-05	1.91E-92	5.80E-3
Illnesses of father: Alzheimer's disease/dementia	0	0.221	1.08E-121	0.465

Illnesses of mother: Alzheimer's disease/dementia	0	0.411	5.63E-117	0.661
Illnesses of siblings: Alzheimer's disease/dementia	0	0.789	4.30E-114	0.816
Enrichment for ATC anti-Parkinson（N04）				
Irritability	0	0.156	4.34E-12	0.770123035
Worrier / anxious feelings	**0**	**6.37E-2**	**3.80E-14**	**2.07E-2**
Illnesses of father: Alzheimer's disease/dementia	**6.51E-101**	**0.459**	**1.07E-13**	**0.106**
Illnesses of mother: Alzheimer's disease/dementia	3.21E-299	0.860	9.35E-15	0.851
Illnesses of siblings: Alzheimer's disease/dementia	5.47E-39	0.417	4.88E-14	0.553
Enrichment for ATC antipsychotics（N05A）				
Irritability	0	0.223	1.58E-20	0.604
Worrier / anxious feelings	0	8.83E-2	8.63E-20	0.832
Illnesses of father: Alzheimer's disease/dementia	7.59E-188	0.780	1.89E-24	0.614
Illnesses of mother: Alzheimer's disease/dementia	0	0.950	3.33E-24	0.996
Illnesses of siblings: Alzheimer's disease/dementia	1.44E-67	0.617	2.83E-23	0.631
Enrichment for ATC anxiolytics or antidepressants（N05B or N06A）				
Irritability	0	1.01E-06	6.30E-112	0.257
Worrier / anxious feelings	**0**	**8.11E-3**	**3.06E-101**	**9.76E-2**
Illnesses of father: Alzheimer's disease/dementia	0	0.155	9.54E-141	0.471
Illnesses of mother: Alzheimer's disease/dementia	0	0.711	1.24E-145	0.356
Illnesses of siblings: Alzheimer's disease/dementia	0	0.967	1.49E-136	0.961
Enrichment for all ATC psychiatric drugs（N05 or N06）				
Irritability	0	0.885	8.60E-146	1.00

Worrier / anxious feelings	0	0.949	5.97E-134	1.00
Illnesses of father: Alzheimer's disease/dementia	0	0.792	1.72E-179	0.980
Illnesses of mother: Alzheimer's disease/dementia	0	0.928	2.28E-186	1.00
Illnesses of siblings: Alzheimer's disease/dementia	0	0.701	4.64E-176	0.829

Drug enrichment analysis based on the definition of MEDI-HPS

In the enrichment analysis conducted on drug sets defined by MEDI-HPS, we further explored the roles of drugs associated with specific mental illnesses (such as depression and anxiety disorders) in the treatment of Alzheimer's Disease (AD) (Supplementary Table S14). The results indicated that drugs related to depression and anxiety disorders exhibited highly significant enrichment across multiple AD-related traits. For example, in traits associated with a parental history of AD/dementia, antidepressant drugs showed higher significance, suggesting that these medications may positively impact the emotional symptoms of AD patients by regulating mood and improving cognitive function.

Drug Enrichment Analysis Based on ClinicalTrials

Furthermore, the drug enrichment analysis based on clinical trial data revealed that drugs associated with irritability and Alzheimer's Disease (AD) were significantly enriched in multiple tissues, whereas drugs related to anxiety/worry exhibited relatively lower enrichment (Supplementary Table S15). This result suggests that certain medications may be more suitable for intervening in specific AD symptoms, while other drugs may demonstrate better efficacy for different symptoms.

Overall, the results of the drug enrichment analysis provide valuable insights into how to select the most appropriate medications to address various AD symptoms. In particular, the significant enrichment of psychotropic drugs in emotion management indicates that they may play a crucial role in the personalized treatment of AD. Future research should further validate the specific mechanisms of these drugs and evaluate their efficacy and safety in AD patients through clinical trials, thereby offering more comprehensive and effective treatment options for patients.

Discussion

Implications for Alzheimer's Disease Pathogenesis

In this study, by integrating multi-tissue Transcriptome-Wide Association Studies (TWAS) and multimodal deep learning methods, we identified several key genes associated with Alzheimer's Disease (AD). These genes play significant roles in the pathogenesis of AD, particularly in the regulation of emotions, cognition, and behavior. The APOE gene, as a core risk gene for AD, significantly influences disease progression through its ε4 allele by regulating cholesterol metabolism, lipid transport, and amyloid plaque clearance. The joint action of APOC1 and APOE also plays a crucial role in lipid metabolism and immune regulation and is significantly associated with emotional symptoms (such as anxiety and irritability), suggesting its potential impact on the emotional and behavioral regulation in AD. The SCIMP gene is expressed in immune cells and participates in immune signal transduction, further highlighting the importance of immune responses in AD-related neuroinflammation.

Additionally, the KAT8 gene encodes a histone acetyltransferase involved in gene expression regulation. Its mutations are associated with neurodevelopmental disorders, indicating its dual role in the emotional and behavioral regulation mechanisms of AD. Other genes, such as RP11-196G11.6 and VKORC1, are involved in epigenetic regulation and vitamin K metabolism, respectively, suggesting that AD may involve multiple pathological pathways, including vascular pathology and cognitive impairment. PRSS53 and ZNF646 are associated with peripheral tissue lesions, particularly exhibiting regulatory potential in metabolic and inflammatory pathways. Mitochondrial dysfunction-related gene TOMM40 and cell adhesion-related gene NECTIN2 further indicate that AD may not only be a neurodegenerative disease but also involve disruptions in peripheral metabolism and the immune system.

The discovery of these genes underscores the multifaceted pathological mechanisms of AD, suggesting that its pathology is not confined to the central nervous system. Pathological changes in peripheral tissues such as the liver and blood vessels may also play important roles in disease progression. Gene Ontology (GO) and KEGG pathway analyses further validated the central roles of pathways like cholesterol metabolism and immune responses in AD-related phenotypes, especially highlighted by the significant enrichment results in brain regions such as the anterior cingulate cortex. These findings

provide new perspectives for understanding the complex pathological mechanisms of AD and offer potential targets for future personalized treatments.

Performance and interpretability of AD-MIF

In this study, the proposed AD-MIF model demonstrated significant advantages in integrating multimodal data to predict Alzheimer's Disease (AD)-related tasks. AD-MIF enhances the model's generalization capability and prediction accuracy by combining raw omics features with omics neighborhood features and efficiently extracting and integrating these features using autoencoders and graph autoencoders. Compared to traditional machine learning methods (such as SVM, LR, RF), deep learning methods (such as ResNet, RNN), and modality imputation-based deep learning methods (such as DeepGAMI, GenNet), AD-MIF achieved higher AUC scores in predicting multiple AD-related tasks, particularly excelling in handling high-dimensional and sparse data. This indicates that AD-MIF can effectively capture complex feature information, supporting early screening and personalized treatment of AD.

To improve the model's interpretability and uncover the biological significance of genes in prediction tasks, the AD-MIF model incorporates an attention mechanism. By calculating the attention weights of each gene using Gen-G, the model better understands their relative contributions to different phenotype prediction tasks. The gene importance score analysis in Figure 6 reveals that genes related to cognitive level include ADRA1A, CAMKK1, and TRAPPC4, which are involved in neural transmission, regulation of cognitive functions, and protein transport; genes associated with CERAD scores include SLC30A4, KLHL5, and UCHL1, primarily involved in zinc ion transport and protein degradation; key genes for Braak stages such as PML, WASF2, and CNTN3 are mainly related to neuroprotection and synapse formation; and genes related to the APOE genotype, including CSRNP2, OXTR, and IRAK4, are involved in the regulation of social behaviors and inflammatory responses. Notably, POLR2C and TRAPPC4 exhibit high associations across multiple phenotypes, suggesting their potentially central roles in the pathogenesis of Alzheimer's Disease (AD). POLR2C, as a subunit of RNA polymerase II, regulates gene expression related to neural functions and neuroprotection. TRAPPC4 plays a crucial role in protein transport and vesicle formation, maintaining protein homeostasis and synaptic transmission within neurons. The multi-phenotype associations of these two genes

suggest that gene expression regulation and protein metabolism may act synergistically in AD, deepening our understanding of the complex molecular mechanisms of AD.

AD-MIF not only showcases exceptional predictive capabilities in Alzheimer's Disease research but also serves as a powerful tool for understanding the disease's multifaceted pathological mechanisms. Through gene importance score analysis, the model identifies key candidate genes associated with immunity, metabolism, and neuroinflammation, providing directions for future research. These findings can further validate the biological functions of these genes and explore their potential as therapeutic targets.

Complementary Roles of PrediXcan and AD-MIF in Phenotypic Prediction

In this study, based on the ROSMAP dataset, after predicting the cognitive diagnosis phenotype, we performed enrichment analyses of the relevant important genes using both individual-level PrediXcan and the deep multi-layer information fusion model AD-MIF. The results revealed significant overlaps between the two methods in the following key pathways and tissue categories: GO Component: GO:0005622 - Intracellular anatomical structure, as well as multiple tissue categories including BTO:0000042 - Animal, BTO:0000142 - Brain, BTO:0001484 - Nervous system, and BTO:0001489 - Whole body. These overlaps indicate that PrediXcan and AD-MIF exhibit high consistency in identifying genes associated with intracellular anatomical structures and multiple key tissue systems, particularly highlighting the enrichment of genes in the brain and nervous system. This underscores the importance of these genes in the regulation of cognitive functions.

Furthermore, the enrichment analysis revealed the unique strengths of each method, demonstrating their complementarity. Individual-level PrediXcan showed significant gene enrichment in multiple organelle-related GO components (such as GO:0043231, GO:0005829, GO:0043229, GO:0043227, GO:0043226) as well as in UniProt Keywords including “Alternative splicing” (KW-0025), “Phosphoprotein” (KW-0597), and “Acetylation” (KW-0007). This highlights PrediXcan’s unique advantages in gene expression regulation and protein modification processes. On the other hand, AD-MIF demonstrated significant gene enrichment in more refined cellular substructure categories such as GO:0005654 - Nucleoplasm and GO:0031981 - Nuclear lumen. This

suggests that AD-MIF's multi-layered information fusion capability can capture more nonlinear and complex gene interactions, identifying key genes and pathways that PrediXcan might not detect.

This complementarity indicates that the combined use of PrediXcan and AD-MIF can more comprehensively elucidate the molecular mechanisms underlying cognitive diagnosis phenotypes. By fully leveraging the strengths of each method, the research depth and breadth are enhanced, providing a more holistic understanding of the molecular underpinnings of cognitive functions in Alzheimer's Disease.

Drug Repositioning Potential

In this study, through drug repositioning analysis, we identified a series of drugs that may have therapeutic potential for Alzheimer's Disease (AD). For example, Sirolimus and Dasatinib demonstrated significant efficacy across multiple AD-related phenotypes. Sirolimus may slow neuroinflammation by modulating immune responses, while Dasatinib shows therapeutic potential by clearing senescent cells, thereby alleviating neurodegenerative changes and associated emotional symptoms. These drugs may mitigate disease progression, cognitive decline, and emotional symptoms by specifically regulating the expression of AD-related genes.

The identification of drugs such as Nintedanib, Quizartinib, and MK-2206 further suggests the potential of non-traditional medications in AD treatment. These drugs may exert neuroprotective effects through mechanisms such as reducing inflammatory responses, inhibiting RIPK1 signaling, and regulating cell survival. Although these medications have not been widely used for AD treatment, their successful application in other indications provides a basis for their potential in AD.

The drug repositioning results of this study emphasize that personalized drug interventions based on specific genetic backgrounds may be more effective. Future research should further validate the effects of these drugs in in vivo and in vitro models to explore their applicability in AD treatment.

Limitations

Although this study achieved significant findings by integrating multi-tissue Transcriptome-Wide Association Studies (TWAS) and deep learning techniques, it also

has several limitations. Firstly, while the TWAS method can reveal associations between gene expression and diseases, it cannot directly establish causality. Therefore, experimental validation is required to further confirm the functional impacts of these genes. Secondly, the effectiveness of the multimodal fusion algorithm depends on the quality of the input data. Future research needs to incorporate more high-quality multi-omics data to validate the findings. Additionally, the results of the drug repositioning analysis also need to be further validated through clinical trials to ensure the efficacy and safety of these drugs in practical applications.

Future Directions

Future research can further explore the following aspects: (1) **Experimental Validation of Significant Genes:** Conduct additional experimental validations to elucidate the functions of significant genes in the pathogenesis of Alzheimer's Disease (AD). This will help establish causal relationships and better understand the biological mechanisms underlying AD. (2) **Expansion of Multimodal Fusion Algorithms:** Enhance the multimodal fusion algorithms by integrating more types of omics data, such as metabolomics and proteomics. This integration will enable a more comprehensive uncovering of the complex pathological mechanisms of AD, providing deeper insights into the disease's multifaceted nature. (3) **Large-Scale Clinical Trials for Drug Efficacy:** Validate the efficacy of potential therapeutic drugs through large-scale clinical trials. Particular emphasis should be placed on exploring the interactions between these drugs and key genes, such as APOE, to ensure targeted and effective interventions. This step is crucial for translating preclinical findings into safe and effective clinical applications for AD patients.By addressing these areas, future studies will enhance the understanding of AD's molecular mechanisms and contribute to the development of more effective and personalized treatment strategies.

Conclusions

This study, through multi-tissue Transcriptome-Wide Association Studies (TWAS) and multimodal deep learning methods, unveiled the complex genetic mechanisms of Alzheimer's Disease (AD), emphasizing the critical roles of multiple genes in pathological processes such as cholesterol metabolism, immune responses, and neuroinflammation. Notably, genes like APOE and APOC1 are particularly prominent in regulating emotions, cognition, and behavior. Additionally, the research findings

indicate that AD is not only a neurodegenerative disease but may also involve pathological changes in peripheral tissues. Through drug repositioning analysis, several drugs with potential therapeutic effects were identified, providing new directions for the precision treatment of AD. Future studies should continue to explore the clinical applications of these key genes and potential drugs to advance the development of personalized treatments.

In this study, the AD-MIF model also played a significant role by integrating multimodal data, thereby enhancing the feature extraction and prediction capabilities for AD-related genes. The application of this model not only improved the prediction accuracy across multiple AD-related tasks but also provided robust support for formulating personalized treatment plans. Future research should further optimize the AD-MIF model and continuously validate the efficacy of these key genes and potential drugs in clinical settings to achieve more precise intervention strategies.

Methods

Data Sources and Preprocessing

GWAS Summary Statistics: This study utilized GWAS summary statistics obtained by Jiang et al., who analyzed UK Biobank data using a generalized linear mixed model (GLMM) method, such as fastGWA-GLMM. These data underwent rigorous quality control and phenotype standardization, covering 61 traits in areas including mental disorders, neurological diseases, sleep patterns, and metabolic conditions, with all samples of European ancestry. The sample sizes ranged from hundreds to hundreds of thousands of individuals, ensuring the breadth and statistical significance of the analysis (see Appendix 1 for detailed information). The data preprocessing involved the following main steps:

1. **Coordinate Conversion:** Due to differences between versions of the genome reference sequences, the GWAS data were first converted from human genome version 19 (hg19) to the latest reference version (hg38) to ensure data consistency and compatibility with modern genomic resources.

2. **Data Imputation:** To handle missing values in the GWAS data, genotype data from the 1000 Genomes reference panel were used in combination with imputation tools. We performed imputation of missing variants using a 100kb sliding window

approach and utilized standardized dosage information to enhance the accuracy of the imputation.

3. **Variant Filtering:** Only variant sites that met quality control standards were retained, including variants with a minor allele frequency (MAF) > 0.01 and a genotype call rate > 0.8, to ensure high data quality and reliability.

4. **Ancestry Matching:** To reduce biases due to population structure, the data analysis was restricted to samples of European ancestry.

This series of processing steps ensured the high quality of the GWAS data, provided a reliable foundation for subsequent analyses, and helped enhance the signal detection capabilities related to the target phenotypes.

GTEx Dataset: This study utilized gene expression prediction models for 49 different tissues from the GTEx dataset. These models were based on eQTL models constructed using the mashr method in the GTEx dataset, and each tissue-specific model can provide predictions of gene expression levels in the corresponding tissue.

ROSMAP Dataset: This study utilized data from the Religious Orders Study and Memory and Aging Project (ROSMAP), managed by the Rush Alzheimer's Disease Center in Chicago. ROSMAP is a longitudinal clinical-pathological study specifically focused on Alzheimer's Disease (AD) and its related pathologies. The dataset includes genotype data for 22 autosomes and gene expression data from the dorsolateral prefrontal cortex, ensuring that each subject's genotype and gene expression data are matched for use. Additionally, we collected detailed phenotype data, including sex, education level, ethnicity, APOE genotype, age at death, MMSE scores, pathological scores, and cognitive diagnoses. These phenotype data provide rich background information for gene-phenotype association analyses, aiding in a deeper understanding of the functions of AD-related genes.

In the data preprocessing phase, we performed the following steps:

1. **Genotype Data Filtering:** Exclude sites that are not single nucleotide variants (SNVs) or have multiple alternate alleles to ensure data quality.

2. **SNP Matrix Conversion:** Convert the filtered genotype data into an SNP matrix for subsequent association analyses.

3. **SNP Annotation:** Annotate SNPs using the UCSC hg19 reference genome and the

dbSNP137 database, extracting the RSIDs for each SNP and adding them to the information table.

4. **Gene Annotation:** Obtain gene chromosomal locations, gene IDs, and types from the Ensembl database, and filter out genes on sex chromosomes to ensure only autosomal gene annotation data are retained.

All data underwent stringent quality control, and the analyses were based on resources provided by the AD Knowledge Portal.

Drug-induced gene expression profile data: The drug-induced gene expression profile data were sourced from the LINCS L1000 project, which is part of the Connectivity Map (CMap) dataset, focusing on the gene expression impacts of drugs and small molecule compounds. LINCS L1000 utilizes a gene expression measurement technique that captures expression values for 978 landmark genes and infers the expression of the entire genome through modeling. These data are used to analyze the gene expression effects of drugs to explore potential molecular mechanisms and for drug repositioning studies.

In the data preprocessing phase, we performed the following steps:

1. **Data Filtering and Parsing:** From the raw data, select sample IDs with treatment type labeled as "trt_cp" (drug treatment).

2. **Expression Statistics Parsing:** Use the cmapR package to parse the data, extracting and saving the expression statistics matrix for subsequent analyses.

Clinical trial drugs and indications data： To validate the effectiveness of the drug repositioning results, we utilized data from three drug sources: (1) ATC Classification Drugs Provided by KEGG: This includes antiepileptic, antiparkinsonian, psychotropic, antipsychotic, and antidepressant/anxiolytic drugs. (2) MEDI-HPS: A dataset containing high-precision validated drug indication information. (3) ClinicalTrials.gov: Clinical trial drug lists provided by ClinicalTrials.gov. These data provided the foundation for drug enrichment testing and were used to combine p-values for different tissues using Fisher and Tippett methods.

Association between genetic prediction gene expression level and 61 brain-related characteristics

In this study, we employed the S-PrediXcan method using GWAS summary statistics provided by Jiang et al., combined with data from the UK Biobank, to investigate the associations between 61 brain-related traits and genetically predicted gene expression levels. These GWAS summary statistics were generated using a generalized linear mixed model (GLMM) approach, such as fastGWA-GLMM. We applied gene expression prediction models from 49 different tissues in the GTEx dataset to these data and inferred gene expression levels in each tissue using the S-PrediXcan method.

For each tissue, we calculated the associations between predicted gene expression and trait risks using the prediction models, generating corresponding p-values. Based on these results, we conducted visual analyses of correlations and pleiotropy findings. We performed multiple comparison corrections on the p-values from the Transcriptome-Wide Association Studies (TWAS) of the 61 brain traits across the 49 tissues, generating corresponding q-values to identify significantly associated gene expression data.

These significant gene expression associations were then visualized using various heatmaps and normalized significance count methods to better understand the pleiotropy and correlations of genes across different brain traits. This comprehensive approach allowed us to elucidate the complex relationships between genetically predicted gene expression and a wide range of brain phenotypes, providing deeper insights into the multifaceted genetic architecture underlying Alzheimer's Disease and related cognitive functions.

Training of gene expression prediction model and database construction

We utilized genotype data from 553 samples of autosomal chromosomes and gene expression data from the dorsolateral prefrontal cortex within the ROSMAP project to train gene expression prediction models. For this purpose, we adopted the PredictDB method to convert the trained models into a format compatible with the PrediXcan software suite. Following the GTEx protocol, and given that the sample size exceeded 350, we incorporated 60 PEER factors into the models. The model training steps

included:

1. **Data Preparation:** Calculated residuals based on PEER factors, utilizing gene expression files and covariate files for data preparation.
2. **Model Training and Database Construction:** Trained models using the PredictDB method and stored the results in an SQLite database for subsequent analyses.

Model Training and PrediXcan Analysis at Individual Level

After training the gene expression prediction models and constructing the database, we employed the individual-level PrediXcan method to predict gene expression levels based on these models and evaluated their associations with target phenotypes using the available genotype datasets. The analysis steps for individual-level PrediXcan included:

1. **Gene Expression Prediction:** Using the trained models, we predicted the gene expression levels of individuals in the cohort.
2. **Phenotype Association Analysis:** We performed association analyses between the predicted gene expression and target phenotypes (such as APOE genotype, age at death, pathological scores, and cognitive diagnoses) to assess the impact of gene expression on these traits.

Deep Multi-layer Information Fusion Model AD-MIF

In response to the multi-modal feature data obtained from ROSMAP samples, this study further proposes a deep multi-layer information fusion model, AD-MIF. This method first preprocesses the AD genotype data to alleviate the challenges posed by the high dimensionality and feature sparsity of genotype data. During the data preprocessing phase, the experiment constructs a multi-branch multilayer perceptron to efficiently extract the genotype data features $\mathrm{G}_{\mathrm{i}}^{\mathrm{P}}$:

$$\mathrm{G}_{\mathrm{i}}^{\mathrm{P}} = MLP_i\left(\mathrm{G}_{\mathrm{i}}^{\widehat{\mathrm{P}}}\right) \tag{1}$$

Where $\mathrm{G}_{\mathrm{i}}^{\widehat{\mathrm{P}}}$ denotes the initial genotype data on the i-th chromosome, and MLP_i represents the multilayer perceptron corresponding to the i-th chromosome.

High-dimensional features lead to a surge in the number of parameters in deep learning models, while sparsity limits the model's ability to mine low-order correlations between features, thereby affecting prediction accuracy. For each type of modality data, considering the multiple interaction relationships between neighborhoods, AD-MIF uses Euclidean distance to identify the top K most similar feature data for each gene or SNP as its neighborhood. Based on this, a K-NN gene/SNP neighborhood feature graph representation (Gen-G / SNP-G) is constructed. Subsequently, AD-MIF employs Autoencoders (AE) and Graph Autoencoders (GAE) to extract latent feature information from the original omics features and the omics neighborhood feature graphs, respectively. To integrate these latent feature representations, AD-MIF combines the feature information through a linear combination to obtain a more comprehensive latent representation:

$$Z_I^i = \alpha Z_{EC}^i + (1 - \alpha) Z_{GEC}^i \tag{2}$$

Where α denotes the learnable feature weight, initialized to 0.5.

Finally, to better utilize both the original omics features and the neighborhood features, AD-MIF introduces an information fusion module. This module aims to integrate the specific and shared features learned from the autoencoders at both local and global levels, thereby generating highly consistent feature representations. The local and global feature representations are combined through element-wise summation to obtain the final fused representation:

$$Z^i = Z_L^i + \beta Z_G^i \tag{3}$$

Where Z_L^i and Z_L^i represent the local and global feature representations, respectively, and β is the feature weight parameter.

The model accomplishes various downstream AD phenotype prediction tasks to validate the application value of AD-MIF in multiple AD-related tasks. The architecture of AD-MIF is illustrated in Figure 10, and the detailed experimental procedures can be found in the Supplementary Methods. AD-MIF employs deep learning techniques to perform deep fusion of multimodal data, extracting more representative features to enhance prediction accuracy.

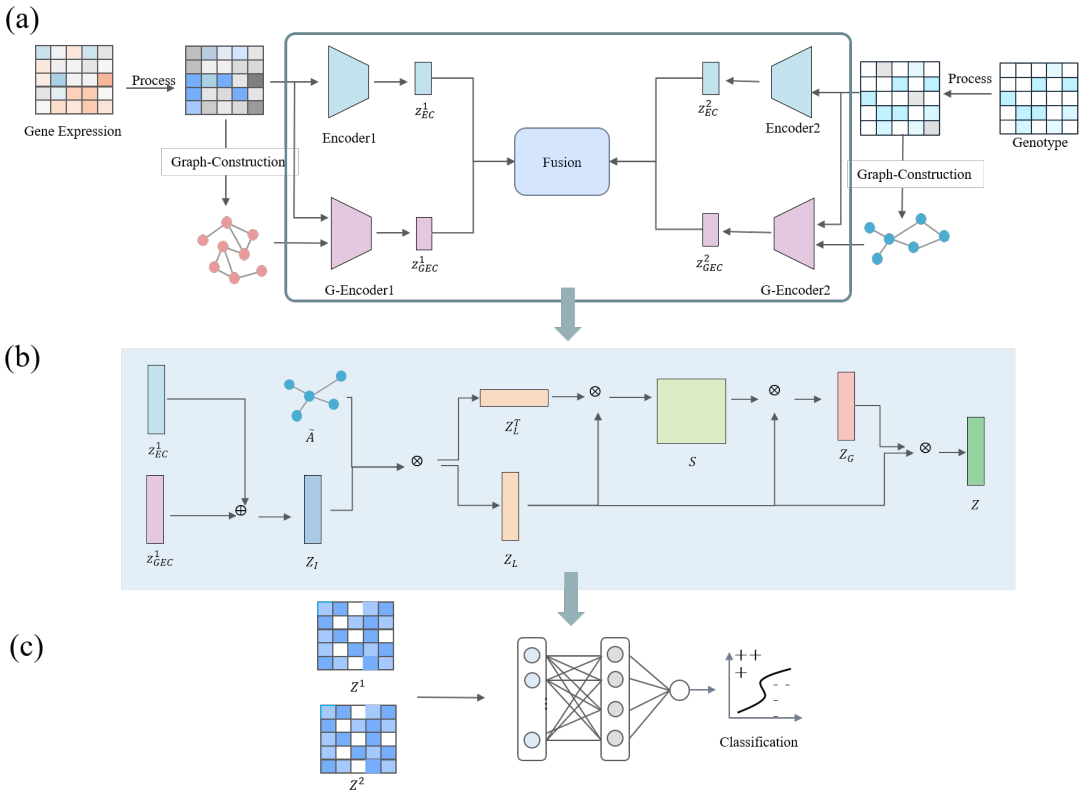


Figure 10. AD-MIF architecture diagram. **a)** Initialization of gene features and SNP features, and the construction process of neighborhood feature graphs; **b)** Information fusion process (using genes as an example); **c)** Prediction based on different downstream AD tasks.

Gene-based and Gene-set Analyses

Based on the correlations and pleiotropic findings displayed in the charts, we selected features with a richer number of significant genes from the S-PrediXcan results (including irritability, worry/anxiety, father's disease: Alzheimer's disease/dementia, mother's disease: Alzheimer's disease/dementia, and siblings' disease: Alzheimer's disease/dementia). Additionally, combining the individual-level PrediXcan analysis results using ROSMAP data, we conducted enrichment analyses on these significant genes. Through these analyses, we further investigated how genes regulate the occurrence and development of these traits in different tissues. The specific methods are as follows:

1. **Gene-Level Enrichment Analysis of S-PrediXcan Results:** For each trait, enrichment analyses were performed across multiple tissues. By integrating gene expression levels and q-values, we identified which genes have significant

regulatory roles in specific tissues.

2. **Enrichment Analysis of AD Family History Traits:** For traits related to the family history of Alzheimer's disease (such as the history of Alzheimer's disease or dementia in fathers, mothers, and siblings), similar enrichment analyses were conducted to further examine the distribution of significant genes across different tissues.

3. **Enrichment Analysis of Individual-Level PrediXcan Phenotype Association Results:** Based on the AD-related phenotype association analysis results from the ROSMAP data, enrichment analyses were performed to identify significant genes and regulatory pathways associated with Alzheimer's disease (AD). These analysis results help us better understand the potential roles of genes in the pathogenesis of AD and provide a foundation for subsequent functional validations and the discovery of potential drug targets.

Drug Repositioning via Gene Expression Correlation

This study is based on the method developed by So et al., which explores potential drug repositioning candidates by comparing disease-related gene expression changes with drug-induced gene expression profiles. We utilized Spearman and Pearson correlation coefficients to assess the correlation between gene expression profiles inferred from GWAS summary statistics and drug-induced expression profiles provided by CMap. Specifically, the Spearman correlation coefficient was used to evaluate the rank-order correlation between the two variables, while the Pearson correlation coefficient was employed to assess the linear relationship between them. Subsequently, we performed the Kolmogorov-Smirnov test to determine whether the genes that are upregulated and downregulated in the disease state exhibit similar expression changes under drug treatment. This test compares the differences between two empirical cumulative distribution functions to evaluate whether the drug can reverse the disease-related gene expression patterns.

Additionally, we conducted drug repositioning analysis on the significant genes identified in the phenotype association results of the APOE genotype. By comparing the correlation between the gene expression profiles associated with the APOE genotype and the drug-induced expression profiles, we further screened for candidate drugs that may have therapeutic potential for APOE-related pathological processes.

This series of analytical procedures ultimately helped us uncover potential therapeutic drugs for AD from multiple dimensions, providing a foundation for drug repositioning and the precision treatment of the disease.

Drug Set Repositioning Enrichment Analysis Method

To evaluate the enrichment of known drug sets in transcriptome-wide association analyses, we employed two primary testing methods:

1. **Self-Contained Test:** This test assesses the significance of specific drug sets. Specifically, we used a one-sample t-test to determine whether the z-scores of the target drug sets significantly differ from zero.
2. **Competitive Test:** This test evaluates significance by comparing the rankings of drugs within the set to those outside the set. We performed a two-sample t-test to compare differences between the target drug sets and randomly selected drug sets.

For control analyses, we randomly selected an equal number of drugs from the dataset to generate random drug sets. For each disease and brain region dataset, we conducted one-sided t-tests and converted the p-values to z-scores using the Probit function. Subsequently, we performed meta-analyses on the p-values from different tissues using the Fisher method and Tippett's minimum p-value method to integrate results from multiple independent experiments. These approaches assess the significance of drug sets by combining multiple p-values. To control for multiple testing issues, we applied False Discovery Rate (FDR) correction using the Benjamini-Hochberg method. A significance threshold of 0.05 was set to ensure the reliability of the analysis results.

This analytical workflow enables us to effectively evaluate the potential therapeutic roles of drug sets in Alzheimer's disease and other complex disease studies.

Data availability

The GWAS summary statistics for 2,989 binary traits analyzed by Jiang et al. using UKB data can be accessed at http://fastgwa.info/ukbimpbin and the GWAS Catalog (GCP ID: GCP000224). The GTEx data is publicly available through the GTEx Portal (https://gtexportal.org). Gene expression, genotype, and phenotypic data from the dorsolateral prefrontal cortex were provided by the Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago. Data collection was supported by National Institute on Aging (NIA) grants P30AG10161 (ROS), R01AG15819 (ROSMAP;

genomics and RNAseq), R01AG17917 (MAP), and U01AG46152 (ROSMAP AMP-AD). Additional phenotypic data can be requested at www.radc.rush.edu. The LINCS L1000 Level 5 data is accessible via NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70138). The MEDI resource can be found at https://www.vumc.org/cpm/center-precision-medicine-blog/medi-ensemble-medication-indication-resource. A list of drugs included in clinical trials (according to ClinicalTrials.gov) is available at https://doi.org/10.15363/thinklab.d212. The Anatomical Therapeutic Chemical (ATC) Classification from KEGG can be accessed via http://www.genome.jp/kegg-bin/get_htext?br08303.keg. Lastly, the MetaXcan tool is available at https://github.com/hakyimlab/MetaXcan.

The key results of this study, including gene expression predictions, significant trait associations, and drug repositioning outcomes, are available at https://drive.google.com/drive/folders/1K6qg5YY2nGCXbXfzBr5hg6aKWwowsd2e.

Code availability

The code used for the analyses in this study is available at the following GitHub repository:https://github.com/LearnDrugAI/Multi-Tissue_Analysis_DL_AD.

Acknowledgements

We are deeply grateful to the patients and their families for their generous contributions to this research. This study made use of data obtained from the AD Knowledge Portal (https://adknowledgeportal.org), specifically gene expression, genotype, and phenotypic data from the dorsolateral prefrontal cortex, provided by the Rush Alzheimer's Disease Center at Rush University Medical Center, Chicago. Data collection was supported by grants from the National Institute on Aging (NIA), including P30AG10161, R01AG15819, R01AG17917, and U01AG46152. We also acknowledge the vital role of research volunteers and the valuable contributions of collaborating researchers. Additionally, this work was supported by the scientific and technological project "Research and Development of Key Technologies for the Production and Quality Control of Sipaiyi Intelligent Preparations," led by Lv Xiaoyi in 2023.

Ethics declarations

Competing interests

The authors declare no competing interests.

References

[1] Scheltens P, De Strooper B, Kivipelto M, Holstege H, Chételat G, Teunissen CE, et al. (2021). Alzheimer's disease. Lancet, https://doi.org/10.1016/s0140-6736(20)32205-4.

[2] James SL, Theadom A, Ellenbogen RG, Bannick MS, Montjoy-Venning W, Lucchesi LR, et al. (2019). Global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology, https://doi.org/10.1016/S1474-4422(18)30415-0.

[3] Kumar. A, Sidhu. J, Lui. F, Tsao. JW. Alzheimer Disease. 2024 Jan-. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. Available from: https://www.ncbi.nlm.nih.gov/books/NBK499922/.

[4] Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. (2015). Alzheimer's disease. Nature Reviews Disease Primers, https://doi.org/10.1038/nrdp.2015.56.

[5] (2024). 2024 Alzheimer's disease facts and figures. Alzheimers Dement, https://doi.org/10.1002/alz.13809.

[6] Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. The American Journal of Human Genetics, https://doi.org/10.1016/j.ajhg.2017.06.005.

[7] Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. Nature Genetics, https://doi.org/10.1038/s41588-022-01024-z.

[8] Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat Genet, https://doi.org/10.1038/ng.3367.

[9] Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nature Genetics, https://doi.org/10.1038/s41588-019-0385-z.

[10]Parrish RL, Buchman AS, Tasaki S, Wang Y, Avey D, Xu J, et al. (2024). SR-TWAS: leveraging multiple reference panels to improve transcriptome-wide association study power by ensemble machine learning. Nature Communications, https://doi.org/10.1038/s41467-024-50983-w.

[11]Guo S, Yang J. (2024). Bayesian genome-wide TWAS with reference transcriptomic data of brain and blood tissues identified 141 risk genes for Alzheimer's disease dementia. Alzheimer's Research & Therapy, https://doi.org/10.1186/s13195-024-01488-7.

[12]Mews MA, Naj AC, Griswold AJ, Below JE, Bush WS. (2024). Brain and Blood Transcriptome-Wide Association Studies Identify Five Novel Genes Associated with Alzheimer's Disease. medRxiv, https://doi.org/10.1101/2024.04.17.24305737.

[13]Consortium TG, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, et al. (2020).

The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science, https://doi.org/doi:10.1126/science.aaz1776.

[14] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med, https://doi.org/10.1371/journal.pmed.1001779.

[15] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, https://doi.org/10.1038/s41586-018-0579-z.

[16] De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, et al. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Scientific Data, https://doi.org/10.1038/sdata.2018.142.

[17] Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nature Communications, https://doi.org/10.1038/s41467-018-03621-1.

[18] Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nature Genetics, https://doi.org/10.1038/ng.3367.

[19] Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. (2019). A guide to deep learning in healthcare. Nature Medicine, https://doi.org/10.1038/s41591-018-0316-z.

[20] van Hilten A, Kushner SA, Kayser M, Ikram MA, Adams HHH, Klaver CCW, et al. (2021). GenNet framework: interpretable deep learning for predicting phenotypes from genetic data. Communications Biology, https://doi.org/10.1038/s42003-021-02622-z.

[21] Chandrashekar PB, Alatkar S, Wang J, Hoffman GE, He C, Jin T, et al. (2023). DeepGAMI: deep biologically guided auxiliary learning for multimodal integration and imputation to improve genotype–phenotype prediction. Genome Medicine, https://doi.org/10.1186/s13073-023-01248-6.

[22] van der Flier WM, de Vugt ME, Smets EMA, Blom M, Teunissen CE. (2023). Towards a future where Alzheimer's disease pathology is stopped before the onset of dementia. Nature Aging, https://doi.org/10.1038/s43587-023-00404-2.

[23] Singh B, Day CM, Abdella S, Garg S. (2024). Alzheimer's disease current therapies, novel drug delivery systems and future directions for better disease management. Journal of Controlled Release, https://doi.org/https://doi.org/10.1016/j.jconrel.2024.01.047.

[24] Xia Y, Sun M, Huang H, Jin W-L. (2024). Drug repurposing for cancer therapy. Signal Transduction and Targeted Therapy, https://doi.org/10.1038/s41392-024-01808-1.

[25] Zhan P, Yu B, Ouyang L. (2022). Drug repurposing: An effective strategy to accelerate contemporary drug discovery. Drug Discov Today, https://doi.org/10.1016/j.drudis.2022.05.026.

[26] So H-C, Chau CK-L, Chiu W-T, Ho K-S, Lo C-P, Yim SH-Y, Sham P-C. (2017). Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. Nature Neuroscience, https://doi.org/10.1038/nn.4618.

[27] Jiang L, Zheng Z, Fang H, Yang J. (2021). A generalized linear mixed model association tool for biobank-scale data. Nature Genetics,

https://doi.org/10.1038/s41588-021-00954-4.

[28] Parolo S, Mariotti F, Bora P, Carboni L, Domenici E. (2023). Single-cell-led drug repurposing for Alzheimer's disease. Scientific Reports, https://doi.org/10.1038/s41598-023-27420-x.

[29] Huang F, Liang J, Lin Y, Chen Y, Hu F, Feng J, et al. (2023). Repurposing of Ibrutinib and Quizartinib as potent inhibitors of necroptosis. Communications Biology, https://doi.org/10.1038/s42003-023-05353-5.

[30] Wang Y, Che L, Chen X, He Z, Song D, Yuan Y, Liu C. (2023). Repurpose dasatinib and quercetin: Targeting senescent cells ameliorates postmenopausal osteoporosis and rejuvenates bone regeneration. Bioactive Materials, https://doi.org/https://doi.org/10.1016/j.bioactmat.2023.01.009.

[31] Lauterbach EC. (2016). Repurposing psychiatric medicines to target activated microglia in anxious mild cognitive impairment and early Parkinson's disease. Am J Neurodegener Dis.

[32] Sills GJ. (2023). Pharmacological diversity amongst approved and emerging antiseizure medications for the treatment of developmental and epileptic encephalopathies. Therapeutic Advances in Neurological Disorders, https://doi.org/10.1177/17562864231191000.

[33] Quiñonez-Flores A, Martinez-Guerra BA, Román-Montes CM, Tamez-Torres KM, González-Lara MF, Ponce-de-León A, Rajme-López S. (2024). Cephalotin Versus Dicloxacillin for the Treatment of Methicillin-Susceptible Staphylococcus aureus Bacteraemia: A Retrospective Cohort Study. Antibiotics.

[34] Tuan NM, Lee CH. (2019). Penfluridol as a Candidate of Drug Repurposing for Anticancer Agent. Molecules, https://doi.org/10.3390/molecules24203659.

[35] Petrich AM, Leshchenko V, Kuo P-Y, Xia B, Thirukonda VK, Ulahannan N, et al. (2012). Akt inhibitors MK-2206 and nelfinavir overcome mTOR inhibitor resistance in diffuse large B-cell lymphoma. Clinical Cancer Research.

[36] Murdin L, Hussain K, Schilder AGM. (2016). Betahistine for symptoms of vertigo. Cochrane Database of Systematic Reviews, https://doi.org/10.1002/14651858.CD010696.pub2.

[37] Hayat M, Gao T, Cao Y, Rafiq M, Zhuo L, Li Y-Z. (2024). Docking Study of Licensed Non-Viral Drugs to Obtain Ebola Virus Inhibitors. International Journal of Pharmaceutical Research and Allied Sciences.

[38] El-Remaily MAEAAA, Aboelez MO, Ezelarab HAA, Selim HMRM, Taha EA, Mohamed SK, et al. (2024). Guanidine dicycloamine-based analogs: green chemistry synthesis, biological investigation, and molecular docking studies as promising antibacterial and antiglycation leads. Molecular Diversity, https://doi.org/10.1007/s11030-024-10816-w.

[39] Choi H-J, Jeong YJ, Kim J, Hoe H-S. (2023). EGFR is a potential dual molecular target for cancer and Alzheimer's disease. Frontiers in Pharmacology.

[40] Huang Y, Ma M, Zhu X, Li M, Guo M, Liu P, et al. (2021). Effectiveness of idebenone nanorod formulations in the treatment of Alzheimer's disease. Journal of Controlled Release, https://doi.org/https://doi.org/10.1016/j.jconrel.2021.06.024.

[41] Zang C, Zhang H, Xu J, Zhang H, Fouladvand S, Havaldar S, et al. (2023). High-throughput target trial emulation for Alzheimer's disease drug repurposing with real-world data. Nature Communications, https://doi.org/10.1038/s41467-023-43929-1.

[42] Das V, Miller JH, Alladi CG, Annadurai N, De Sanctis JB, Hrubá L, Hajdúch M. (2024). Antineoplastics for treating Alzheimer's disease and dementia: Evidence from

preclinical and observational studies. Medicinal Research Reviews.

[43] Munafò A, Burgaletto C, Di Benedetto G, Di Mauro M, Di Mauro R, Bernardini R, Cantarella G. (2020). Repositioning of immunomodulators: a ray of hope for Alzheimer's disease? Frontiers in Neuroscience.

[44] Kang YJ, Diep YN, Tran M, Cho H. (2020). Therapeutic targeting strategies for early-to late-staged Alzheimer's disease. International Journal of Molecular Sciences.