

Article

A Comprehensive Evaluation of Cross-Omics Blood-Based Biomarkers for Neuropsychiatric Disorders

Weichen Song ¹, Weidi Wang ¹, Zhe Liu ¹, Wenxiang Cai ¹, Shunying Yu ², Min Zhao ²
and Guan Ning Lin ^{1,*}

¹ School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China; goubegou@sjtu.edu.cn (W.S.); wwd-swxx@foxmail.com (W.W.); liuzm@sjtu.edu.cn (Z.L.); caiwenxiang@sjtu.edu.cn (W.C.)

² Shanghai Key Laboratory of Psychotic Disorders, Shanghai 200030, China; yushuny@yahoo.com (S.Y.); drzhaomin@gmail.com (M.Z.)

* Correspondence: nicknlin@sjtu.edu.cn

Abstract: The identification of peripheral multi-omics biomarkers of brain disorders has long been hindered by insufficient sample size and confounder influence. This study aimed to compare biomarker potential for different molecules and diseases. We leveraged summary statistics of five blood quantitative trait loci studies (N = 1980 to 22,609) and genome-wide association studies (N = 9725 to 500,199) from 14 different brain disorders, such as Schizophrenia (SCZ) and Alzheimer's Disease (AD). We applied summary-based and two-sample Mendelian Randomization to estimate the associations between blood molecules and brain disorders. We identified 524 RNA, 807 methylation sites, 29 proteins, seven cytokines, and 22 metabolites having a significant association with at least one of 14 brain disorders. Simulation analyses indicated that a cross-omics combination of biomarkers had better performance for most disorders, and different disorders could associate with different omics. We identified an 11-methylation-site model for SCZ diagnosis (Area Under Curve, AUC = 0.74) by analyzing selected candidate markers in published datasets (total N = 6098). Moreover, we constructed an 18-methylation-sites model that could predict the prognosis of elders with mild cognitive impairment (hazard ratio = 2.32). We provided an association landscape between blood cross-omic biomarkers and 14 brain disorders as well as a suggestion guide for future clinical discovery and application.

Keywords: peripheral biomarker; neuropsychiatric disorder; mendelian randomization; diagnostic model; prognosis



Citation: Song, W.; Wang, W.; Liu, Z.; Cai, W.; Yu, S.; Zhao, M.; Lin, G.N. A Comprehensive Evaluation of Cross-Omics Blood-Based Biomarkers for Neuropsychiatric Disorders. *J. Pers. Med.* **2021**, *11*, 1247. <https://doi.org/10.3390/jpm11121247>

Academic Editors: Naoya Oribe and Takefumi Ueno

Received: 18 October 2021

Accepted: 23 November 2021

Published: 24 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The diagnosis of chronic brain disorders at the present day is primarily dependent on clinical symptom assessments, which suffers from the drawback of subjectivity, symptom heterogeneity, and disease comorbidity [1]. To overcome these difficulties and aid early intervention of brain disorders, researchers have made considerable efforts to find objective diagnostic and predictive biomarkers [2]. Among all potential biomarkers, peripheral blood molecules, such as RNA [3], methylation site [4], and proteins [5] have gained specific attention due to the high feasibility and relatively low costs. So far, researchers have found a large number of potential blood biomarkers for the diagnosis of brain disorders [6].

However, the study and application of blood-based biomarkers has long suffered from the lack of reproducibility [6]. Overfitting is one of the main challenges: a transcriptome-wide analysis typically has more than 20,000 RNAs detected, whereas the sample size of test subjects is usually limited to no more than a few hundred due to the labor cost burden. The conflicts between feature number and the sample size are even exacerbated when the biomarkers' discovery strategy of "multi-omics" is applied [7]. On the other hand, technical, and biological confounders, such as batch effect, immune state, lifestyle, and influence

from diseases of different organs, dramatically impact the level of blood molecules [6], which further beclouds the true association between biomarkers and brain disorders.

Theoretically, Mendelian Randomization (MR) [8] could properly overcome the challenge of overfitting and confounder effects (Figure S1). By using genetic variants that significantly impact the expression level of a blood molecule (so-called Quantitative Trait Loci, QTL) [9–13] as instruments, MR separates participants into high- and low-level groups. Next, MR compares the prevalence of a disorder between two groups to evaluate the association between the molecule and the disorder. Furthermore, in a 2-sample MR (2SMR) scenario, the QTL-molecule and QTL-disease associations can be estimated in two different Genome-Wide Association Studies (GWAS) (Figure S1) and then integrated to estimate the final molecule-disease association. Such analyses overcome overfitting by profoundly increasing the effective sample size (i.e., combining samples of two GWAS) and eliminating confounders' impact since the genetic instruments are unlikely to be influenced by environmental factors [8]. Based on the framework of 2SMR, Zhu et al. [14] proposed a new MR method, Summary statistic-based Mendelian Randomization (SMR), which better addresses the MR analysis of high-throughput data. Both SMR and 2SMR can serve as the ideal tools for biomarker discovery.

The explosive growth of large-scale QTL studies demonstrated the significant heritability of blood levels of RNA [11], methylation [12], protein [13], cytokine [10], and metabolite [9], which provided the opportunity of applying 2SMR and SMR to discover their association with brain disorders. In the current study, we integrated these QTL data and GWAS summary statistics of 14 brain disorders (Figure S1) to achieve three goals: (1) comprehensively evaluate the potential of biomarkers for each omic and disorder; (2) validate some of the top candidates using publicly available high-throughput datasets; and (3) provide a biomarker potentiality landscape, which could serve as a reference for a future biomarker study.

2. Materials and Methods

2.1. Data Collection and Pre-Processing

We downloaded blood QTL [9–13] and disease GWAS summary statistics from the public domain, including SCZ [15], Bipolar Disorder (BP) [16], Major Depressive Disorder (MD) [17], Anorexia Nervosa (AN) [18], Attention deficit hyperactivity disorder (ADHD) [19], General Anxiety Disorder (ANX) [20], Tourette's syndrome (TS) [21], Obsessive-Compulsive Disorder (OCD) [22], Autism Spectrum Disorder (ASD) [23], Alcohol Dependence (ALD) [24], Post-traumatic stress disorder (PTSD) [25], Alzheimer's Disease (AD) [26], Parkinson's Disease (PD) [27] and Nicotine Dependence (ND) [28]. We searched for all blood QTL and GWAS studies that (1) provided effect size and standard error summary statistics for genome-wide variants; (2) recruited only European populations; (3) for the QTL study, only involved healthy participants. For each omic and disorder, we chose the study with the largest sample size. Readers who wish to replicate the study should gain approval from the corresponding authors of these data. We applied uniform filtration, pre-processing, and Surrogate Variable Analysis [29] (SVA)-based confounder adjustment on all datasets (Supplementary Notes).

2.2. Summary-Based and 2-Sample Mendelian Randomization (SMR and 2SMR)

For the association between RNA, methylation, protein markers, and brain disorders, we applied multi-SNP-based SMR [30], which utilized all cis-QTLs ($p < 5 \times 10^{-8}$) within 1-MB window of the markers to estimate the association. For the association between metabolite, cytokine markers, and brain disorders, we applied 2SMR by R package TwoSampleMR [31] using SNP with $p < 1 \times 10^{-5}$ on the entire genome. For blood-based markers of cytokines and metabolites, their QTLs did not have spatial implications, i.e., no "cis" or "trans" QTLs of a cytokine or a metabolite. Therefore, instead of SMR, we applied classic 2SMR, which included genome-wide significant QTLs as instruments, regardless of

their genomic positions. For each QTL i of a marker m , SMR and 2SMR first estimated the effect of m on a disease d ($\beta_{md(i)}$) by Wald ratio

$$\beta_{md(i)} = \frac{\beta_{id}}{\beta_{im}}$$

where β_{id} denoted effect of i on d (i.e., GWAS effect size of i) and β_{im} denoted effect of i on m (i.e., QTL effect size of i). The SE (and corresponding statistics z) for each QTL was estimated by the delta method

$$SE_{md(i)} = \frac{SE_{id}}{\beta_{im}}$$

SMR and 2SMR then applied different methods to integrate multiple QTL results into the final estimation and corresponding p -value (Supplementary Notes). p -value adjustment was conducted separately for each omic-disease combination. Inflation factor λ was calculated on a quantile-quantile plot for each omic disease combination. We defined λ as the slope of Chi-square regression of actual p -value on expected p -value [32].

Next, we applied the Heterogeneity In Dependent Instruments test (HEIDI) [14] to evaluate whether these associations were driven by the co-localization (i.e., molecule level and disease do not share the same causal SNP, but their causal SNPs were in a strong Linkage disequilibrium). If the association was driven by the same causal SNP instead of colocalized SNPs, the HEIDI test would be expected to return a p -value > 0.05 , for which we denoted the marker as HEIDI(+), p -value > 0.05 would be denoted as HEIDI(−).

2.3. Simulation Analysis

To quantify the classification power of markers from each omics, we generated simulation data with the hypothesis that SMR-estimated β_{md} truly reflected reality, and with the consideration of estimation uncertainty and environmental influence. Specifically, for each omic-disease combination, we repeated the following procedure 1000 times to generate 1000 simulation datasets:

(1) For marker m ($m = 1, 2, \dots, n$) from omic o of disease d , we generated normal distribution $\beta_{md} \sim N(\beta_{md}, SE_{md})$, where β_{md} and SE_{md} were effect size and SE obtained from SMR or 2SMR. We then generated a random β_{md} from the normal distribution, which formed an effect size vector $B_{od} = \{\beta_{md}\}_{m=1,2,\dots,n}$.

(2) We then generated a random expression matrix $E_{10,000 \times n}$ by generating n random vectors of length 10,000 from $N(0,1)$. This was because all OR from GWAS or QTL analysis has been standardized, such that β_{md} corresponded to log odds of d per 1-SD increment of m . To account for environmental confounders, we added a random noise of $N(0,0.01)$ on each vector.

(3) We calculated the odds of d as $ODD(d) = \{odd_i\}_{i=1,2,\dots,10,000} = E \times B_{od}$, and subsequently, the probability of d as $P(d) = \{p_i\}_{i=1,2,\dots,10,000} = \left\{ \frac{1}{1+odd_i} \right\}_{i=1,2,\dots,10,000}$. For simplicity, the intercept term was set as zero, i.e., the number of cases of d is set to be identical to that of control.

(4) The label (case or control) for each of the 10,000 simulated samples was randomly decided, with the probability of being a case = $P(d)$.

On each of the simulation datasets, we applied Logistic regression by rms R package, and recorded the AUC and R2. We took the median AUC and R2 across 1000 simulation for comparison. For cross-omic analysis, we pooled all markers of a disease, ranked them according to the absolute effect size, and generated simulation datasets of all these markers by the same procedure. In each simulation data, we sequentially applied Logistic regression on top 1, top2, . . . top n markers and recorded the AUC, R2, and AIC (by MASS R package). We calculated the median values across 1000 simulations, and chose the optimal model with the lowest median AIC. All the above simulation analysis was carried out separately for HEIDI(+) and HEIDI(−) markers.

2.4. Published Transcriptome and Methylome Data Analysis

For public transcriptome or methylome data, we extracted the value of HEIDI(+) and HEIDI(−) markers, applied Logistic regression, and recorded the AUC. The obtained AUC was compared to the corresponding simulation AUC (restricted to markers available in the real data). To compare the power of HEIDI(+) and HEIDI(−) markers, we ranked the HEIDI(−) markers according to their SMR p -value and chose top markers with the same number of HEIDI(+) markers. We applied Logistic regression on these two sets of markers of the same number and compared their AUC and log-likelihood.

2.5. Diagnostic Model Construction

For SCZ methylation markers, we calculated the Spearman correlation coefficient ρ between each of the 1897 SMR-identified markers (both HEIDI(+) and HEIDI(−)) and diagnostic status, and retained only those with (1) ρ and SMR β of same direction; (2) $|\rho| > 0.05$. Then, we applied a Bayesian LASSO (bLASSO) regression by monomvn R package on the remaining 480 markers. All markers with median posterior coefficients not equal to zero were chosen as candidate marker. In the training set, we applied classical LASSO regression on the candidate markers. All remaining markers, together with their non-zero coefficients, constructed the final diagnostic model. We determined the optimal cut point using cutpointr R package by maximizing Youden's Index. Finally, the coefficient as well as cut point of the identified model were fixed and applied to the validation set

2.6. Predictive Model Construction

We downloaded from ADNI repository all blood methylation data for which the diagnosis at sample collection was "MCI" (mild cognitive impairment), except those recovered from dementia status. ADNI project collected blood samples from elders at MCI status, and we used these methylation data at the beginning of the observation as potential predictors of future conversion risk. The sample information could be downloaded from ADNI repository after application approval. According to whether the participants converted to AD in the entire follow-up period recorded by ADNI, we classified samples in the training set as converter and non-converter. We first carried out Spearman correlation analysis and LASSO regression similar to the diagnostic model. Then, in the validation set, we applied this model to define high conversion risk and low risk group. The hazard ratio as well as its p value was calculated by univariate Cox regression using survival and survminer R package. Details of diagnostic and predictive model construction can be found in the Supplementary Notes.

3. Results

3.1. Identifying All Potential Blood-Based Biomarkers Associated with Brain Disorders

After data filtering (Supplementary Notes), we collected a set of valid QTLs for blood expression levels of 15,052 RNA, 89,910 methylation sites, 669 proteins, 41 cytokines, and 119 metabolites, on which we applied the MR Method. By restricting at the genome-wide significance threshold ($p < 0.05/n$, where n denoted the number of tested molecules of the corresponding omics), we identified 1386 blood-based molecular markers, including 524 RNAs, 807 methylation sites, 29 proteins, seven cytokines, and 19 metabolites, which had a significant association with brain disorders (Tables S1–S5). We then considered them as potential biomarkers of related diseases in the following analyses.

Next, we applied the Heterogeneity In Dependent Instruments test (HEIDI) to evaluate whether these associations were driven by the co-localization (i.e., molecule level and disease do not share the same causal SNP, but their causal SNPs were in a strong Linkage disequilibrium). Among all potential biomarkers identified by SMR, we found that 82% of RNA, 80% of methylation, and 62% of proteins showed a HEIDI $p < 0.05$ (HEIDI(−)), suggesting that co-localization made a significant contribution to the identified association between peripheral molecules and diseases. Since we do not know whether or not co-localization would influence the biomarker predictivity, we analyzed HEIDI(+)

(HEIDI $p > 0.05$) and HEIDI(−) markers separately in the following section to deal with the potential heterogeneity between them.

3.2. RNA and Methylation Levels Showing Strong Associations with SCZ, PD, and AD

To identify the RNA, methylation, and protein biomarkers and their characteristics, we applied cis-QTL SMR and downstream enrichment analysis on their QTL. When restricted on candidate biomarkers with HEIDI $p > 0.05$ (HEIDI(+)), SCZ had the largest number of RNA ($n = 52$) and methylation ($n = 126$) biomarkers, followed by PD ($n = 8$ and $n = 14$), BP ($n = 10$ and $n = 7$), and AD ($n = 10$ and $n = 6$, respectively, see Figure 1A). These diseases also had a large number of HEIDI(−) markers, as shown in Figure S2. This association was not restricted to the top blood molecules with p -value passing the genome-wide significance threshold, as the general p -value distribution of all molecules was inflated compared to the null uniform distribution (inflation factor $\lambda > 1$; Figure 1A,B). Concordantly, λ was even higher for HEIDI(−) molecules for these disorders (Figures 1B and S2). In contrast, TS, OCD, ND, ALD, and PTSD did not have any RNA or methylation candidate markers, and their λ was also lower than 1 (Figure 1A).

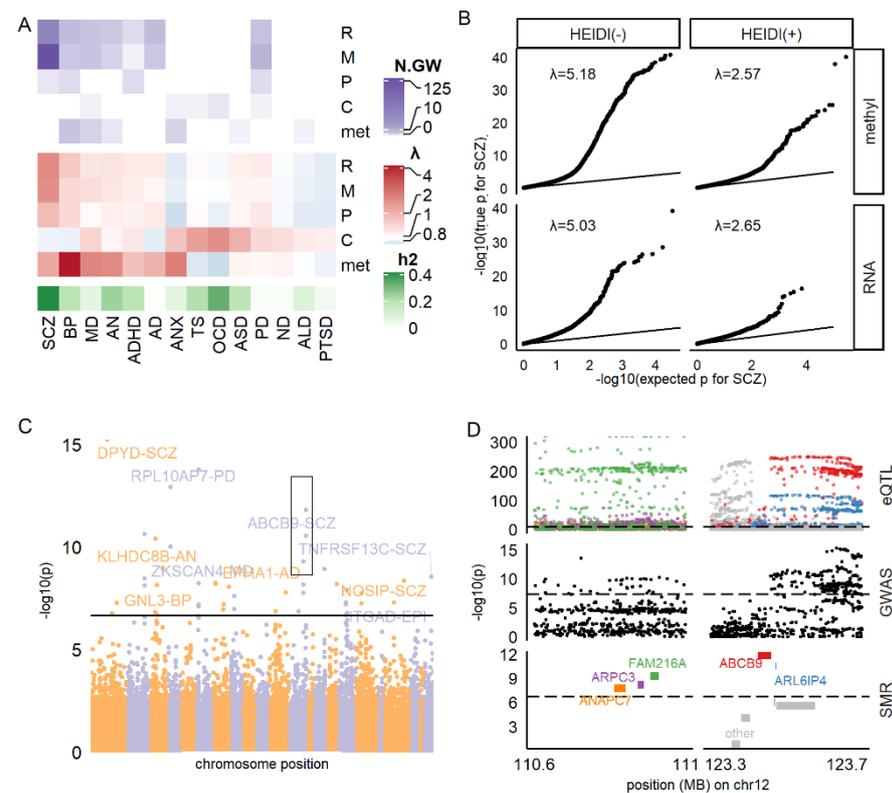


Figure 1. Summary of SMR results. (A) Number of HEIDI(+) molecules reaching genome-wide significance (N.GW; top panel) and inflation factor (middle panel) for each omic and disease. SNP-based heritability (h^2) was estimated by Linkage Disequilibrium Score Regression. (B) QQ plot for SMR result of RNA-SCZ and methylation-SCZ association. (C) Manhattan plot for the association between RNA and all diseases. (D) Regional GWAS, QTL, and SMR statistics for SCZ in distal 12q region (black box in (C)) E: selected top molecules associated with multiple diseases. F: association between cytokines, metabolites, and diseases. R: RNA, m/methyl: methylation site. P: protein. C: cytokine. met: metabolite.

For the RNA-disorder association (Figure 1C), we found 95 HEIDI(+) markers and 429 HEIDI(−) markers reaching the genome-wide significance threshold (Table S1), which gathered on specific regions of the genome. Figure 1D showed an example of a distal 12q region, where five SMR-significant genes (ANAPC7, ARPC3, FAM216A, ABCB9, and ARL6IP4) had close spatial distribution in accordance with the regional QTL and GWAS

loci distribution. However, RNA markers showed neither enrichments in brain-related biological functions (Supplementary Methods; adjusted p -value of Gene Ontology, GO, analysis >0.05) nor enrichment in genes preferentially expressed in the brain [33], or different brain cell types [34] (Fisher's exact test P , $P_{\text{fisher}} > 0.05$; Table S6). We concluded that the identified RNA markers might better reflect global alteration rather than brain abnormality during brain disorders.

For the methylation-disorder association (Figure S3), we found 164 HEIDI(+) and 643 HEIDI(−) markers reaching the genome-wide significance threshold (Table S2). We observed that the methylation markers enriched in genomic regions of DNase I Hypersensitivity Site [35] ($P_{\text{fisher}} = 9.10 \times 10^{-5}$, Odds Ratio [OR] = 1.42), promoters ($P_{\text{fisher}} = 1.99 \times 10^{-5}$, OR = 1.51), prefrontal cortex H3k27ac [36] peaks ($P_{\text{fisher}} = 1.87 \times 10^{-6}$, OR = 1.49), and H3k36me3 [36] peaks ($P_{\text{fisher}} = 2.68 \times 10^{-17}$, OR = 1.88), and the chromHMM [37] state 5 (weak transcription) ($P_{\text{fisher}} = 1.23 \times 10^{-6}$, OR = 1.57, Table S7). These methylation markers tagged 223 proxy genes by falling within the regions, including UTR (17%), gene body (53%), and 1500 base pairs around the transcript start site (30%) of the corresponding genes. Interestingly, these proxy genes enriched in neuronal functions, such as long-term synaptic potentiation (adjusted $P_{\text{fisher}} = 0.002$, Figure S3). They were also nominally enriched in genes preferentially expressed in the brain ($P_{\text{fisher}} = 0.013$, OR = 1.61, Table S6), but not in genes specifically expressed in specific brain cell types. These results suggested that methylation markers might better capture central nervous system abnormality in brain disorders.

We also analyzed protein biomarkers of brain disorders by SMR. We found 11 HEIDI(+) markers reaching significance threshold (Table S3), corresponding to eight different proteins, which included ESAM (PSMR for SCZ = 1.57×10^{-7}), GPNMB (PSMR = 3.01×10^{-6} for PD), FLRT3 (PSMR for BP = 1.06×10^{-5}), and MANBA (PSMR for ADHD = 4.62×10^{-5}). However, due to limited numbers ($n = 11$), they did not show any functional enrichment.

3.3. Cytokines and Metabolites Exhibiting High Pleiotropy

We next analyzed cytokines and metabolites that had an association with brain disorders. As shown in Figure 2 and Table S4, the blood level of IL18 was significantly associated with four disorders: TS ($\beta = 0.19$, $p = 3.69 \times 10^{-7}$), OCD ($\beta = -0.19$, $p = 6.07 \times 10^{-5}$), ANX ($\beta = -0.15$, $p = 0.0002$) and PD ($\beta = 0.10$, $p = 0.0005$). Similarly, another pleiotropic biomarker, the cytokine MIP1B, was also associated with two disorders (ASD: $\beta = 0.07$, $p = 1.33 \times 10^{-6}$; OCD: $\beta = 0.15$, $p = 3.10 \times 10^{-6}$). We also observed the pleiotropic association for metabolite markers, shown in Figure 2A and Table S5. In addition, the ratio of bis-allylic in fatty acid (Bis.FA.ratio) was observed to be significantly associated with two disorders: BP ($\beta = -0.14$, $p = 6.20 \times 10^{-6}$), MD ($\beta = -0.06$, $p = 2.87 \times 10^{-5}$), while multiple metabolites, such as Glycine and ratio of double bonds in fatty acid, were shown to be associated with BP (Figure 2B).

To further illustrate the extent of pleiotropic association between blood-based markers and brain disorders, we summarized the number of associated disorders for all markers (Figure 2A). As stated above, cytokines and metabolites markers exhibited higher pleiotropic associations at the genome-wide significance: IL18, Tyrosine, the bis-allylic ratio in fatty acid, and bis-allylic ratio in double bond were associated with at least three disorders at genome-wide significance. In contrast, for RNA, methylation, and protein markers, only the RNA expression levels of MAPK3 were significantly associated with three disorders (SCZ, PD, and AD). Only when the significance threshold was relaxed to nominal $p < 0.05$, seven methylation biomarkers, such as cg20670488 (Figure 2A), were found associating with eight different disorders. Therefore, in general, at the stringent threshold, cytokines, and metabolites markers exhibited pleiotropic association with brain disorders.

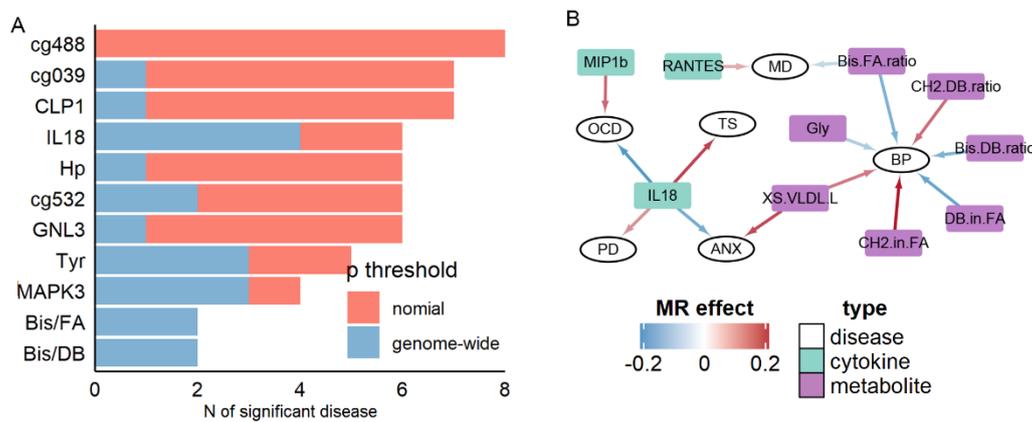


Figure 2. Summary of pleiotropic results. (A) Selected top molecules associated with multiple diseases. (B) Association between cytokines, metabolites, and diseases.

Additionally, we applied step-wise outlier removal and MR sensitivity analysis (Supplementary Notes) to test for causality. We found that the MR result was not explained by pleiotropy or systemic bias (Table S8 and Figure S4), indicating that the association between biomarkers and disorders was causal. Thus, although biomarkers’ discovery did not require them to be causal [5], our potential cytokine and metabolite biomarkers for brain disorders still provided evidence of causality and yielded insights into the disease mechanism.

3.4. Simulation Demonstrated the Advantage of Using Cross-Omics Biomarker Combinations

Having identified the potential biomarkers from each of the five omics, we sought to quantify their diagnostic power and compare them across diseases and omics. Thus, for each omic-disease combination, we generated 1000 simulation datasets based on the estimated effect size of each biomarker (Method) and calculated the Area Under Curve (AUC) and Nagelkerke pseudo-R square (R2) in each dataset.

As shown in Figure 3A, RNA and methylation markers of SCZ and PD had the largest classification power (AUC = 0.78 to 0.96 for HEIDI(+), 0.92 to 0.99 for HEIDI(−)), which were mainly due to a large number of markers (Figure 1A and Figure S2). Nonetheless, some omics had a strong statistical power despite the small number of markers, such as the protein markers of PD (HEIDI(+): 3 markers, AUC = 0.70, R2 = 0.17), RNA markers of AN (HEIDI(+): 5 markers, AUC = 0.85, R2 = 0.46). In sum, we observed that RNA markers generally had the largest classification power: the median AUC for RNA was 0.73, whereas none of the other omics had a median AUC > 0.65. For six diseases (SCZ, AN, BP, MD, ADHD, and AD; Figure 3A), RNA HEIDI(+) markers had a larger AUC than other omics.

We further explored whether a combination of different omics would gain a better diagnostic power. Figure 2B,C showed the examples of BP biomarker combination, which showed association with four omics. We generated simulation data (Method) of all 26 BP markers and calculated the Akaike information criterion (AIC) to choose the optimal combination of markers. The combination with the lowest AIC had the best trade-off between marker numbers and diagnostic efficiency [38]. We found that the model of the top 22 markers had the lowest AIC (Table S9), which achieved the AUC of 0.79 and R2 of 0.33. This model included RNA markers such as GLT8D1, methylation markers such as cg14470998, protein markers such as LMAN2L, and metabolite markers such as CH2.in.FA (CH2 ratio in fatty acid), suggesting that combining markers from different omics might gain a better power of classifying BP patients from a healthy control. A similar analysis was also carried out for HEIDI(−) BP markers, where an 18-marker model consisted of three omics had the lowest AIC (Figure 3C).

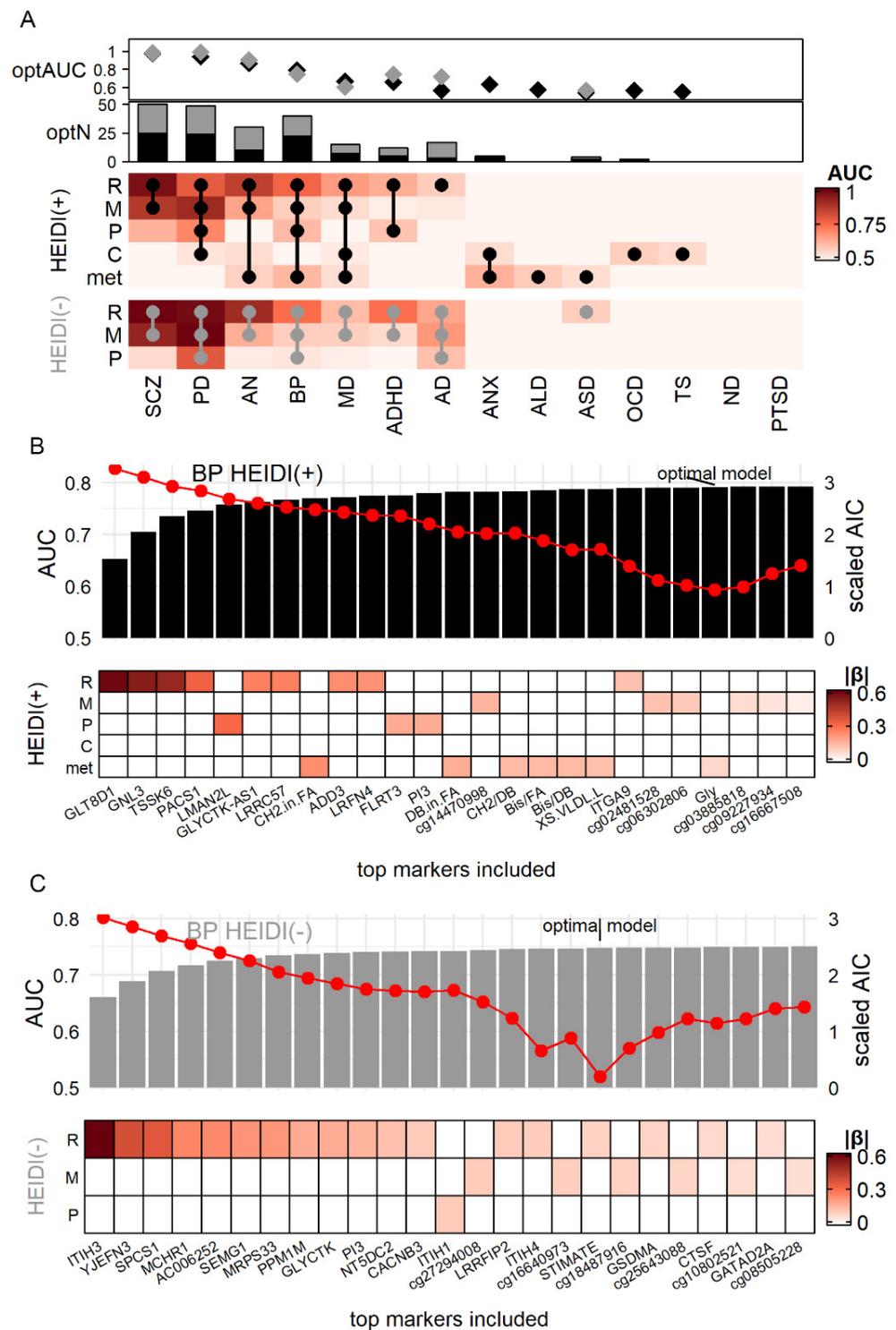


Figure 3. Comparison of biomarker efficiency among diseases and omics in simulation. **(A)** Each grid of the heatmap showed the Area Under Curve (AUC) of the model consisting of all markers from one omic for one disease. Dots on the heatmap showed the composition of the optimal model with the lowest Akaike information criterion (AIC). Diamonds (optAUC: AUC for optimal model) and bar plots (optN: number of markers in optimal model) showed the AUC and number of markers of the optimal models (black for HEIDI(+) model, grey for HEIDI(−) model). **(B,C)** example optimal model of BP. Red dots denoted AIC. R: RNA. M: methylation site. p: protein. **(C)** cytokine. met: metabolite. $|\beta|$: absolute value of β SMR.

Taking the results of all diseases together, we found that the optimal models of seven diseases, from SCZ to AD in Figure 3A, were cross-omics (round dots on Figure 3A denoted the composition of the optimal model). Of note, the optimal HEIDI(+) model of AN achieved AUC = 0.87 and R2 = 0.50 with only 10 multi-omics markers. This model consisted of large-effect markers IMPDH2 (β SMR = 1.48), CADM1 (β SMR = 0.49) (Figure S5). Another noteworthy disease was AD (Figure S6): its HEIDI(+) model had poor performance (three RNA markers, AUC = 0.57, R2 = 0.02), but the HEIDI(−) model (14 markers from three omics, AUC = 0.72, R2 = 0.19) yielded better performance.

3.5. HEIDI(+) and HEIDI(−) Markers Having Comparable Power in Real-World Validation

We next sought to validate the classification power implicated by simulation analysis in the real-world data. We collected 12 cross-sectional blood RNA data [3,39–49] of seven diseases and 11 blood methylation data [4,50–58] of six diseases to evaluate the efficiency of RNA and methylation markers. We did not analyze protein, cytokine, and metabolite markers since limited public data is available. As shown in Figure S7, methylation markers of AD, BP, MD, and AN generally had higher AUC in real data than in simulation data, especially HEIDI(−) markers of AN (real AUC = 0.85, simulation AUC = 0.63). On the other hand, RNA and methylation markers of SCZ and PD tended to have lower AUC in the real data, suggesting that only a small proportion truly took effect among the large number of SMR-identified markers of SCZ and PD.

We then investigated whether the power of HEIDI(−) markers was comparable to HEIDI(+) markers. We observed that HEIDI(+) and HEIDI(−) markers generally had similar AUC. Despite a few exceptions, including methylation markers of AN (HEIDI(+) AUC = 0.76, HEIDI(−) AUC = 0.65), the difference of AUC of HEIDI(+) and HEIDI(−) markers were generally smaller than 0.05. Concordantly, the Likelihood ratio and the number of significant variables of Logistic regression were also similar for HEIDI(+) and HEIDI(−) markers (Table S10), which suggested that their classification power and significance were similar. Therefore, we no longer distinguished HEIDI(+) and HEIDI(−) markers in the following section.

3.6. Construction of Molecular Diagnostic Models for SCZ and AD with Notable Accuracy

Next, we sought to construct optimal models with a subset of top markers by validating the candidate biomarkers in public datasets. After general consideration of biomarker effect size and available sample size, we decided to analyze methylation markers of SCZ, PD, and AD, as well as RNA markers of BP and SCZ (Figures S8 and S9). For SCZ, we divided blood methylation datasets from Hannon et al. [59] into feature selection set (N = 675), training set (N = 547), and validation set (N = 300). Since the number of candidate methylation SCZ markers (N = 1897; Figure 4A) was extremely large, we applied Spearman correlation analysis and Bayesian LASSO (Method), which removed a total of 1856 markers without concordant coefficients in SMR and the feature selection set. Then, we applied classic LASSO on the remaining 41 candidates in the training set (N = 547) and obtained a linear classification model consisted of 11 methylation sites (Figure 4B) with AUC = 0.72 (95% CI of 0.67–0.76). Next, we fixed the coefficient of each predictor as well as the optimal cut point (obtained by maximizing Youden's Index) [60] and applied the model to the validation set (N = 300). The model achieved AUC = 0.74 (95% CI of 0.69–0.80, Figure 4C) with accuracy of 0.70 (sensitivity = 0.71 and specificity = 0.69). The proxy genes of these markers included GABBR1, which encoded a subunit of gamma-aminobutyric acid receptor, SYNGAP1, which encoded a member of N-methyl-D-aspartate receptor complex, and MOG, which took part in oligodendrocyte myelination (Figure 4D). These results indicated that our 11-site model not only robustly classified SCZ patients from healthy control but also had a biological significance that could provide insight into SCZ pathology.

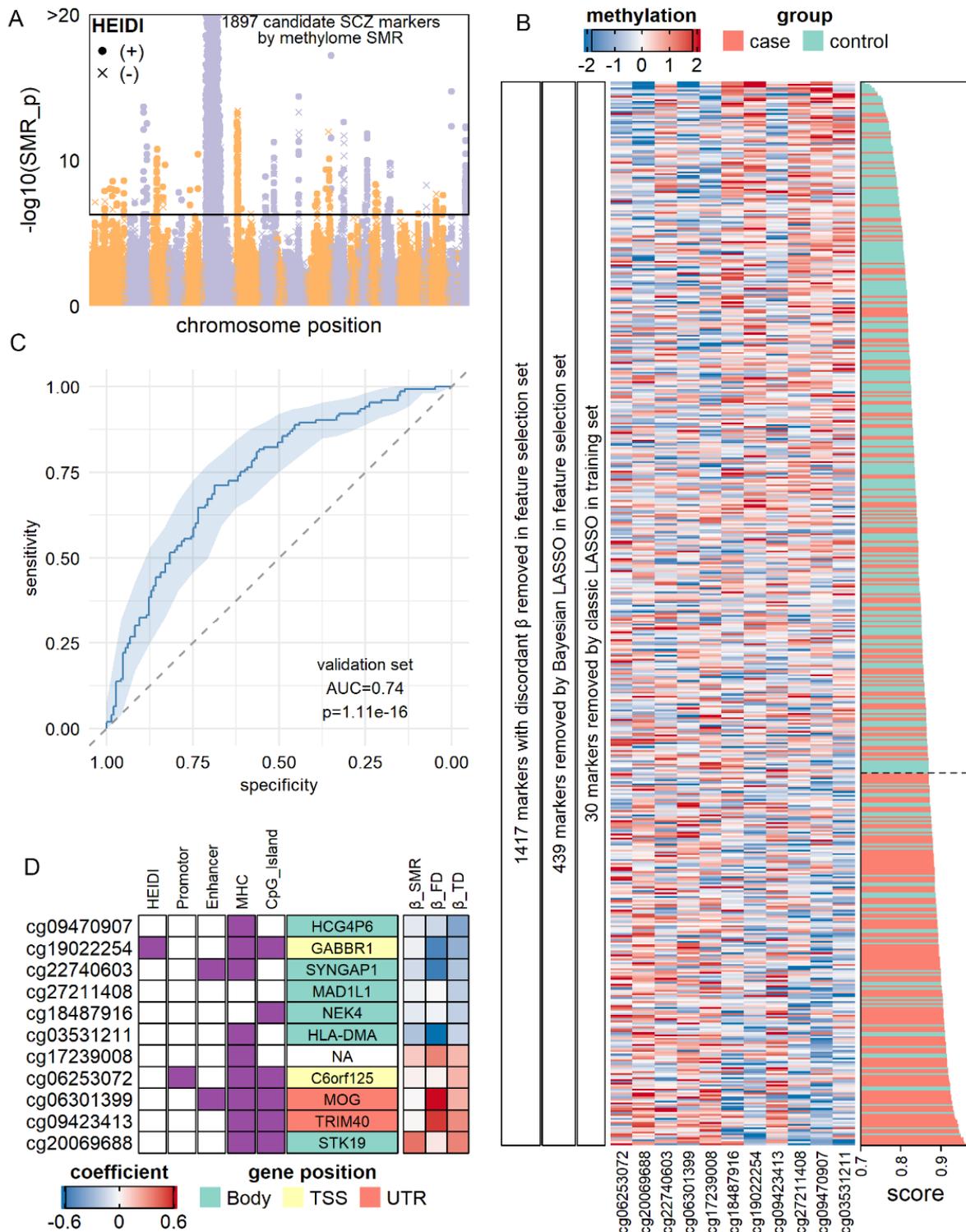


Figure 4. Validation of methylation markers of SCZ. **(A)** Manhattan plot for SCZ-methylation association by SMR. Each dot represents a methylation site. **(B)** LASSO model of SCZ diagnostic model by methylation markers in the training set (N = 547). Heatmap showed the expression levels of 11 markers after filtration and model construction in the training set. Score: diagnostic scores calculated by the sum of methylation levels weighted by the LASSO coefficient. A higher score corresponded to a higher probability of being an SCZ patient. **(C)** Receiver-Operation Curve for LASSO model in the validation set (N = 300). **(D)** Characteristics of methylation markers chosen by LASSO. Gene symbols denoted the proxy genes of each methylation marker, and their grid color denoted the position of the methylation markers on the proxy genes. Purple grids on column “HEIDI” indicated that the markers had $p > 0.05$ in HEIDI. β_{FD} : coefficients in the feature selection set. β_{TD} : coefficients in the training set.

For AD, we downloaded the methylation dataset from ADNI [4] repository. All 74 SMR candidate markers (Figure S10A) were retained for analysis. In the training set (N = 600), we removed 45 out of 74 candidates due to discordant effect size compared with β SMR. We then applied LASSO regression on the remaining 29 candidates and obtained a model of 18 methylation sites (Figure S10B). This model had AUC = 0.79 (95% CI, 0.75–0.83), with diagnostic accuracy = 0.76 (sensitivity = 0.80 and specificity = 0.68). We then applied this model to the validation set (N = 321, Figure S10C) and observed diagnostic AUC = 0.73 (95%CI, 0.67–0.79) and accuracy = 0.70 (sensitivity = 0.73 and specificity = 0.64). These markers recurrently tagged lipoproteins, including, APOE, APOC1, and APOC2 (three times), and APOC4 (Figure S10D), in line with the popular notion that lipoprotein plays a vital role in the pathology of AD [58,61].

We also carried out a similar analysis for PD methylation markers as well as SCZ and BP RNA markers. For PD methylation markers (Figure S8), we obtained a three-site model with AUC = 0.65 (95% CI, 0.62–0.68) in the training set (N = 1200) and AUC = 0.66 (95% CI, 0.61–0.69) in the validation set (N = 689). For BP RNA markers (Figure S9), we found six genes (SPCS1, CTSE, ITGA9, ITIH4, PLAAT3, and PI3) out of 29 candidates in the training set (N = 360). This model achieved AUC = 0.71 (95% CI, 0.67–0.77) in training set and AUC = 0.63 (95% CI, 0.53–0.73) in the validation set (N = 120). For SCZ RNA markers, the identified model did not show significant classification power in the validation set.

3.7. SMR-Identified Methylation Markers Predicting the Risk of AD

As noted, our analysis considered all SMR-identified markers as diagnostic markers that reflected the current status of patients. However, predicting future disease prognosis is also an important task of biomarkers. Thus, we analyzed whether SMR-identified markers could serve as predictive biomarkers. Since a very limited number of the dataset had longitudinal records available, we only analyzed ADNI data of mild cognition impairment (MCI) elders and managed to predict their future conversion to AD by blood methylation data.

As shown in Figure 5A, we analyzed all 74 SMR-identified AD methylation candidates. In the training set (N = 600), LASSO regression returned an 18-site model which could distinguish converters from non-converters at the accuracy of 0.76 (sensitivity, 0.71; specificity, 0.77). The AUC in the training set was 0.79 (95% CI, 0.75–0.83). Using this model and its optimal cut point estimated in the training set, we classified the validation set into a high conversion risk group (N = 102) and a low-risk group (N = 227). In survival analysis (Figure 5B), we found that the high-risk group had a significantly lower interval of non-conversion survival (Hazard ratio = 2.32, $p = 3.1 \times 10^{-5}$). In the low-risk group, 74% (137) of the MCI patients did not convert to AD in the follow-up period of 300 days, whereas, as in the high-risk group, the median non-conversion survival times were 159 days. Similar to the diagnostic model (Figure S10D), these predictive markers' proxy genes, including APOC1, GPC2, and SLC24A4 (Figure 5C). In sum, our results indicated that SMR-identified markers of AD could serve as both diagnostic and predictive markers.

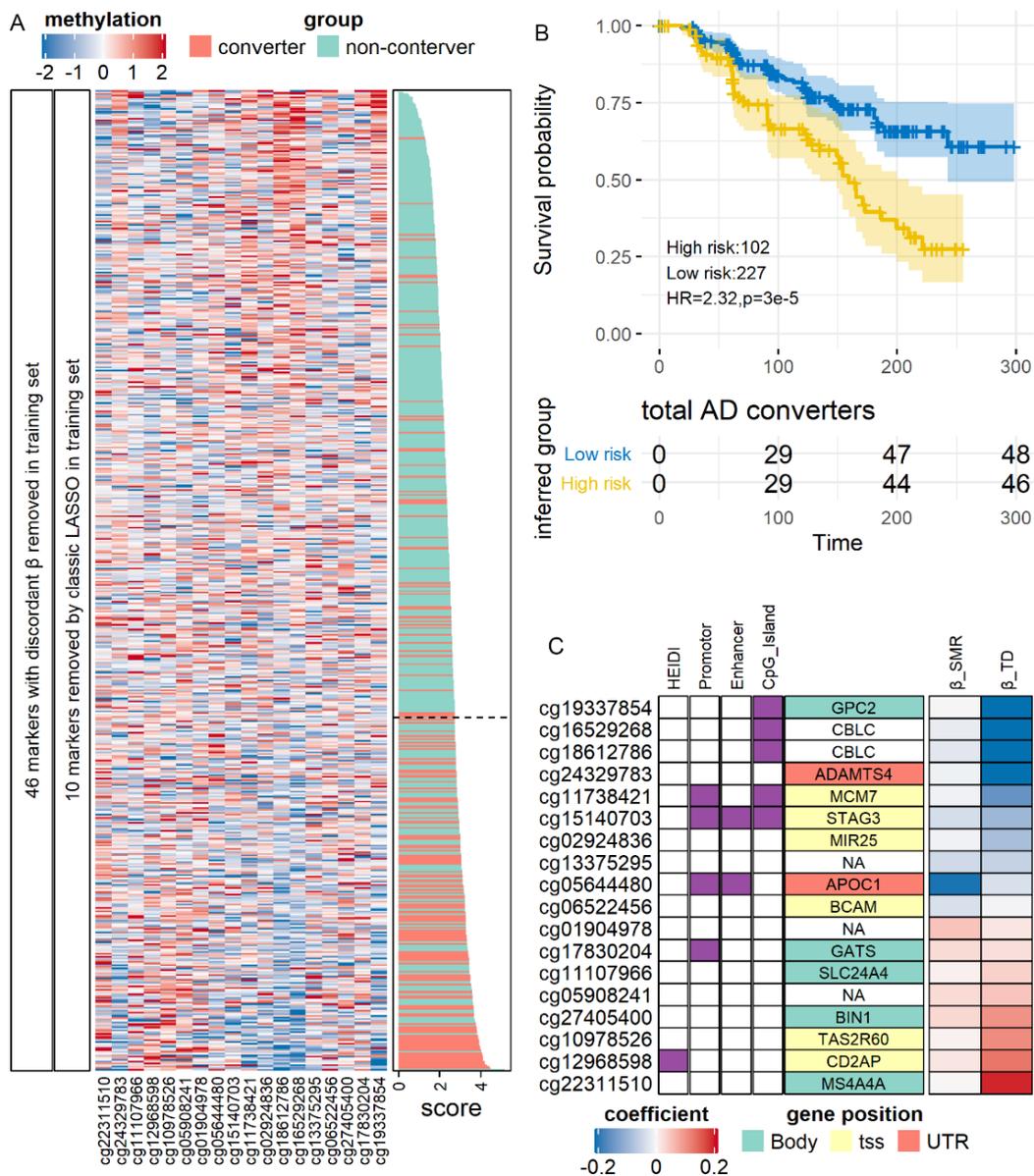


Figure 5. Validation of methylation predictive markers of AD. (A) Expression of markers after filtration and model construction in the training set. (B) Survival curve for conversion-to-AD in the test set. Time was shown in days. (C) Characteristics of the markers. HR: Hazard Ratio.

4. Discussion

In the current study, we applied SMR and 2SMR on the QTL and GWAS statistics to evaluate the association between blood-based molecular markers and different brain disorders. We confirmed that blood levels of various multi-omics molecular markers had a significant association with brain disorders and may serve as both diagnostic and predictive biomarkers.

One of our major findings of the current study is that the peripheral molecules indeed carry information reflecting the central nervous system. The peripheral blood receives substance from all organs and tissues of the body, and its molecular composition is very different from the brain [62]. Traditional cross-sectional analysis has restricted power to decode the potential signals hidden within. In contrast, our MR-based approach, which enabled satisfactory statistical power and confounder-free estimation [8], demonstrated the existence of an association between peripheral signal and brain disorders. Another controversy is that, judging from the biological significance, MR association from colocal-

ized QTL and GWAS SNP is of little interest [14]. However, we found that biomarkers identified by such association (HEIDI(−) markers) generally had comparable power with non-colocalized markers (HEIDI(+) markers). Thus, although co-localization introduced a challenge to GWAS and causality inference, their signals are still valuable for clinical biomarker study. After confirming the significant associations by MR, we further validated a subset of top candidate markers in the public data, and showed that this association was not masked by confounders and could be confidentially detected in the real world. Together, these results showed that the MR-identified candidate markers are highly promising for clinical application.

Furthermore, our result highlighted the importance of multi-omics analysis. Current blood-based biomarker studies were predominantly restricted to transcriptome and methylome, while only a few studies focused on metabolites [63], proteins [5], and cytokines [64]. This discrepancy may be due to the different pace of technology advancement: next-generation sequencing and methylation array could comprehensively quantify genome-wide targets, whereas proteome and metabolome techniques could only cover a proportion of analytes. As a result, researchers might be prone to study transcriptome and methylome. However, our result showed that different brain diseases were associated with different omics, and not one omic could serve as biomarkers for more than half of the involved diseases. Thus, a reasonable solution might be a two-step design: selecting a few promising candidates from multiple omics and applied low-throughput validation in a large cohort. By validation of a few top molecules, we showed that potential markers identified MR in the current study could serve as promising candidates. However, these results could not rule out the impact of medications.

It should be noted that whether a biomarker is diagnostic or predictive is, theoretically, indistinguishable in MR analysis alone. The effect size β of MR is typically interpreted as: (1-SD) increment in biomarker levels changes the odds of disease to $(e\beta)$ -fold [65], similar to the coefficients of Logistic regression. Here, the “odds of disease” could be interpreted in the manners of both diagnosis and prediction: it could be “odds of being a patient now”, or “odds of becoming a patient in the future”. In our analysis of AD methylation markers in ADNI data [4], MR-identified markers fulfill the task of both diagnosis and prediction properly. This result suggested that it is reasonable to interpret the MR-identified markers as either diagnostic or predictive, and it is valuable to validate these markers in both manners.

However, there are still some limitations to the current study. MR requires that the tested molecule has a valid genetic basis. However, many of the blood-based molecules are not controlled by any genetic variants and could not be evaluated by MR. For these molecules, we could not draw conclusion about the lack of biomarker potentiality. Furthermore, there are minimal public data of blood proteome, cytokines, and metabolome available, which prevented us from validating candidate markers using them. The applicability of these candidates should be tested in future real-world studies.

5. Conclusions

In conclusion, our MR revealed that a blood biomarker study should focus on the most promising omics and molecules for the targeted disorders that we highlighted. We summarized all identified biomarkers and highlighted omics in the Supplementary Table S11. Our validation analysis using published data has shown that using this result as reference could profoundly improve the study efficiency and avoid overfitting. Future large multi-omic validation studies could prioritize the highlighted biomarkers in our study to achieve the final clinical models, which would reduce the risk of overfitting and confounder effects.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/jpm11121247/s1>, Supplementary notes, including supplementary methods, supplementary Figures S1–S10 and supplementary Tables S1–S11. Supplementary Table S1–S5 All significant SMR result for transcriptome, methylome, proteome, cytokines, and metabolome. Supplementary Table S6 tissue- and cell type-specific expression. Supplementary Table S7 regional enrichment of methylation

markers. Supplementary Table S8 Sensitivity tests for all heterogeneous MR result. Supplementary Table S9 Simulation analysis for BP. Supplementary Table S10 summary of results of real data analysis. Supplementary Table S11 Summary and comparison of blood multi-omic biomarkers. Supplementary Figure S1 Flowchart and data summary of the study. Supplementary Figure S2 summary of SMR HEIDI(−) results for each omic and disease. Supplementary Figure S3 methylation markers and its biological interpretation. Supplementary Figure S4 Step-wise outlier removal test for Bis.DB.ratio. Supplementary Figure S5 Simulation analysis of AN markers. Supplementary Figure S6 Simulation analysis of AD markers. Supplementary Figure S7 Comparison of HEIDI(+) and HEIDI(−) markers in real-world data. Supplementary Figure S8 Diagnostic model of PD by methylation markers. Supplementary Figure S9 Diagnostic model of BP by RNA markers. Supplementary Figure S10 Validation of methylation diagnostic markers of AD.

Author Contributions: G.N.L. designed the study. W.S. collected the data and performed the analysis. W.W., Z.L. and W.C. helped with the methodology. M.Z. and S.Y. interpreted the results. W.S. and G.N.L. wrote the manuscript. All authors have read, revised, and approved the published version of the manuscript.

Funding: This work was supported by grants from the National Natural Science Foundation of China (No. 81971292 and 81971257); Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (No. 1610000043); Medical-industrial intersection program of Shanghai Jiao Tong University (No. YG2019QNA59 and YG2019ZDA30); Shanghai Mental Health Center (No. 2019-YJ01 and 2019-QH-03).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated in this study can be found in the supplementary tables.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Insel, T.; Cuthbert, B.; Garvey, M.; Heinssen, R.; Pine, D.S.; Quinn, K.; Sanislow, C.; Wang, P. Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **2010**, *167*, 748–751. [[CrossRef](#)] [[PubMed](#)]
- García-Gutiérrez, M.S.; Navarrete, F.; Sala, F.; Gasparyan, A.; Austrich-Olivares, A.; Manzanares, J. Biomarkers in Psychiatry: Concept, Definition, Types and Relevance to the Clinical Reality. *Front. Psychiatry* **2020**, *11*, 432. [[CrossRef](#)] [[PubMed](#)]
- McCaffrey, T.A.; St. Laurent, G.; Shtokalo, D.; Antonets, D.; Vyatkin, Y.; Jones, D.; Battison, E.; Nigg, J.T. Biomarker discovery in attention deficit hyperactivity disorder: RNA sequencing of whole blood in discordant twin and case-controlled cohorts. *BMC Med. Genom.* **2020**, *13*, 160. [[CrossRef](#)] [[PubMed](#)]
- Vasanthakumar, A.; Davis, J.W.; Idler, K.; Waring, J.F.; Asque, E.; Riley-Gillis, B.; Grosskurth, S.; Srivastava, G.; Kim, S.; Kim, S.; et al. Harnessing peripheral DNA methylation differences in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) to reveal novel biomarkers of disease. *Clin. Epigenet.* **2020**, *12*, 84. [[CrossRef](#)]
- Mongan, D.; Föcking, M.; Healy, C.; Susai, S.R.; Heurich, M.; Wynne, K.; Nelson, B.; McGorry, P.D.; Amminger, G.P.; Nordentoft, M.; et al. Development of Proteomic Prediction Models for Transition to Psychotic Disorder in the Clinical High-Risk State and Psychotic Experiences in Adolescence. *JAMA Psychiatry* **2020**, *78*, 77–90. [[CrossRef](#)] [[PubMed](#)]
- Hayashi-Takagi, A.; Vawter, M.P.; Iwamoto, K. Peripheral Biomarkers Revisited: Integrative Profiling of Peripheral Samples for Psychiatric Research. *Biol. Psychiatry* **2014**, *75*, 920–928. [[CrossRef](#)]
- Olivier, M.; Asmis, R.; Hawkins, G.A.; Howard, T.D.; Cox, L.A. The need for multi-omics biomarker signatures in precision medicine. *Int. J. Mol. Sci.* **2019**, *20*, 4781. [[CrossRef](#)] [[PubMed](#)]
- Hemani, G.; Bowden, J.; Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **2018**, *27*, R195–R208. [[CrossRef](#)]
- Kettunen, J.; Demirkan, A.; Würtz, P.; Draisma, H.H.M.; Haller, T.; Rawal, R.; Vaarhorst, A.; Kangas, A.J.; Lyytikäinen, L.P.; Pirinen, M.; et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **2016**, *7*, 11122. [[CrossRef](#)] [[PubMed](#)]
- Ahola-Olli, A.V.; Würtz, P.; Havulinna, A.S.; Aalto, K.; Pitkänen, N.; Lehtimäki, T.; Kähönen, M.; Lyytikäinen, L.P.; Raitoharju, E.; Seppälä, I.; et al. Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am. J. Hum. Genet.* **2017**, *100*, 40–50. [[CrossRef](#)] [[PubMed](#)]
- Võsa, U.; Claringbould, A.; Westra, H.-J.; Bonder, M.J.; Deelen, P.; Zeng, B.; Kirsten, H.; Saha, A.; Kreuzhuber, R.; Kasela, S.; et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* **2018**, *18*, 447367. [[CrossRef](#)]

12. McRae, A.F.; Marioni, R.E.; Shah, S.; Yang, J.; Powell, J.E.; Harris, S.E.; Gibson, J.; Henders, A.K.; Bowdler, L.; Painter, J.N.; et al. Identification of 55,000 Replicated DNA Methylation QTL. *Sci. Rep.* **2018**, *8*, 17605. [[CrossRef](#)] [[PubMed](#)]
13. Sun, B.B.; Maranville, J.C.; Peters, J.E.; Stacey, D.; Staley, J.R.; Blackshaw, J.; Burgess, S.; Jiang, T.; Paige, E.; Surendran, P.; et al. Genomic atlas of the human plasma proteome. *Nature* **2018**, *558*, 73–79. [[CrossRef](#)]
14. Zhu, Z.; Zhang, F.; Hu, H.; Bakshi, A.; Robinson, M.R.; Powell, J.E.; Montgomery, G.W.; Goddard, M.E.; Wray, N.R.; Visscher, P.M.; et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **2016**, *48*, 481–487. [[CrossRef](#)] [[PubMed](#)]
15. Pardiñas, A.F.; Holmans, P.; Pocklington, A.J.; Escott-Price, V.; Ripke, S.; Carrera, N.; Legge, S.E.; Bishop, S.; Cameron, D.; Hamshere, M.L.; et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **2018**, *50*, 381–389. [[CrossRef](#)] [[PubMed](#)]
16. Stahl, E.A.; Breen, G.; Forstner, A.J.; McQuillin, A.; Ripke, S.; Trubetskoy, V.; Mattheisen, M.; Wang, Y.; Coleman, J.R.I.; Gaspar, H.A.; et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **2019**, *51*, 793–803. [[CrossRef](#)]
17. Howard, D.M.; Adams, M.J.; Clarke, T.K.; Hafferty, J.D.; Gibson, J.; Shirali, M.; Coleman, J.R.I.; Hagenaars, S.P.; Ward, J.; Wigmore, E.M.; et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **2019**, *22*, 343–352. [[CrossRef](#)]
18. Watson, H.J.; Yilmaz, Z.; Thornton, L.M.; Hübel, C.; Coleman, J.R.I.; Gaspar, H.A.; Bryois, J.; Hinney, A.; Leppä, V.M.; Mattheisen, M.; et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **2019**, *51*, 1207–1214. [[CrossRef](#)] [[PubMed](#)]
19. Demontis, D.; Walters, R.K.; Martin, J.; Mattheisen, M.; Als, T.D.; Agerbo, E.; Baldursson, G.; Belliveau, R.; Bybjerg-Grauholm, J.; Bækvad-Hansen, M.; et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **2019**, *51*, 63–75. [[CrossRef](#)] [[PubMed](#)]
20. Otowa, T.; Hek, K.; Lee, M.; Byrne, E.M.; Mirza, S.S.; Nivard, M.G.; Bigdeli, T.; Aggen, S.H.; Adkins, D.; Wolen, A.; et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* **2016**, *21*, 1391–1399. [[CrossRef](#)]
21. Yu, D.; Sul, J.H.; Tsetsos, F.; Nawaz, M.S.; Huang, A.Y.; Zelaya, I.; Illmann, C.; Osiecki, L.; Darrow, S.M.; Hirschtritt, M.E.; et al. Interrogating the Genetic Determinants of Tourette’s Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am. J. Psychiatry* **2019**, *176*, 217–227. [[CrossRef](#)]
22. Arnold, P.D.; Askland, K.D.; Barlassina, C.; Bellodi, L.; Bienvenu, O.J.; Black, D.; Bloch, M.; Brentani, H.; Burton, C.L.; Camarena, B.; et al. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **2018**, *23*, 1181–1188. [[CrossRef](#)]
23. Grove, J.; Ripke, S.; Als, T.D.; Mattheisen, M.; Walters, R.K.; Won, H.; Pallesen, J.; Agerbo, E.; Andreassen, O.A.; Anney, R.; et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **2019**, *51*, 431–444. [[CrossRef](#)]
24. Walters, R.K.; Polimanti, R.; Johnson, E.C.; McClintick, J.N.; Adams, M.J.; Adkins, A.E.; Aliev, F.; Bacanu, S.A.; Batzler, A.; Bertelsen, S.; et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **2018**, *21*, 1656–1669. [[CrossRef](#)] [[PubMed](#)]
25. Nievergelt, C.M.; Maihofer, A.X.; Klengel, T.; Atkinson, E.G.; Chen, C.-Y.; Choi, K.W.; Coleman, J.R.I.; Dalvie, S.; Duncan, L.E.; Gelernter, J.; et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat. Commun.* **2019**, *10*, 4558. [[CrossRef](#)] [[PubMed](#)]
26. Jansen, I.E.; Savage, J.E.; Watanabe, K.; Bryois, J.; Williams, D.M.; Steinberg, S.; Sealock, J.; Karlsson, I.K.; Hägg, S.; Athanasiu, L.; et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. *Nat. Genet.* **2019**, *51*, 404–413. [[CrossRef](#)] [[PubMed](#)]
27. Nalls, M.A.; Blauwendraat, C.; Vallerga, C.L.; Heilbron, K.; Bandres-Ciga, S.; Chang, D.; Tan, M.; Kia, D.A.; Noyce, A.J.; Xue, A.; et al. Expanding Parkinson’s disease genetics: Novel risk loci, genomic context, causal insights and heritable risk. *bioRxiv* **2018**, *10*, 388165.
28. Watanabe, K.; Stringer, S.; Frei, O.; Umičević Mirkov, M.; de Leeuw, C.; Polderman, T.J.C.; van der Sluis, S.; Andreassen, O.A.; Neale, B.M.; Posthuma, D. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **2019**, *51*, 1339–1348. [[CrossRef](#)]
29. Leek, J.T.; Johnson, W.E.; Parker, H.S.; Jaffe, A.E.; Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **2012**, *28*, 882–883. [[CrossRef](#)]
30. Wu, Y.; Zeng, J.; Zhang, F.; Zhu, Z.; Qi, T.; Zheng, Z.; Lloyd-Jones, L.R.; Marioni, R.E.; Martin, N.G.; Montgomery, G.W.; et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* **2018**, *9*, 918. [[CrossRef](#)] [[PubMed](#)]
31. Hemani, G.; Zheng, J.; Elsworth, B.; Wade, K.H.; Haberland, V.; Baird, D.; Laurin, C.; Burgess, S.; Bowden, J.; Langdon, R.; et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife* **2018**, *7*, e34408. [[CrossRef](#)] [[PubMed](#)]
32. Aulchenko, Y.S.; Ripke, S.; Isaacs, A.; van Duijn, C.M. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* **2007**, *23*, 1294–1296. [[CrossRef](#)]

33. Genovese, G.; Fromer, M.; Stahl, E.A.; Ruderfer, D.M.; Chambert, K.; Landén, M.; Moran, J.L.; Purcell, S.M.; Sklar, P.; Sullivan, P.F.; et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **2016**, *19*, 1433–1441. [[CrossRef](#)] [[PubMed](#)]
34. Saunders, A.; Macosko, E.Z.; Wysoker, A.; Goldman, M.; Krienen, F.M.; de Rivera, H.; Bien, E.; Baum, M.; Bortolin, L.; Wang, S.; et al. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **2018**, *174*, 1015–1030. [[CrossRef](#)]
35. Skene, N.G.; Grant, S.G.N. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front. Neurosci.* **2016**, *10*, 16. [[CrossRef](#)] [[PubMed](#)]
36. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)] [[PubMed](#)]
37. Andersson, R.; Gebhard, C.; Miguel-Escalada, I.; Hoof, I.; Bornholdt, J.; Boyd, M.; Chen, Y.; Zhao, X.; Schmidl, C.; Suzuki, T.; et al. An atlas of active enhancers across human cell types and tissues. *Nature* **2014**, *507*, 455–461. [[CrossRef](#)] [[PubMed](#)]
38. ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [[CrossRef](#)] [[PubMed](#)]
39. Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J.; Ziller, M.J.; et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317–330. [[CrossRef](#)]
40. Ernst, J.; Kellis, M. ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **2012**, *9*, 215–216. [[CrossRef](#)] [[PubMed](#)]
41. Nho, K.; Nudelman, K.; Allen, M.; Hodges, A.; Kim, S.; Risacher, S.L.; Apostolova, L.G.; Lin, K.; Lunnon, K.; Wang, X.; et al. Genome-wide transcriptome analysis identifies novel dysregulated genes implicated in Alzheimer’s pathology. *Alzheimer’s Dement.* **2020**, *16*, 1213–1223. [[CrossRef](#)] [[PubMed](#)]
42. Krebs, C.E.; Ori, A.P.S.; Vreeker, A.; Wu, T.; Cantor, R.M.; Boks, M.P.M.; Kahn, R.S.; Olde Loohuis, L.M.; Ophoff, R.A. Whole blood transcriptome analysis in bipolar disorder reveals strong lithium effect. *Psychol. Med.* **2019**, *50*, 2575–2586. [[CrossRef](#)]
43. Calligaris, R.; Banica, M.; Roncaglia, P.; Robotti, E.; Finaurini, S.; Vlachouli, C.; Antonutti, L.; Iorio, F.; Carissimo, A.; Cattaruzza, T.; et al. Blood transcriptomics of drug-naïve sporadic Parkinson’s disease patients. *BMC Genom.* **2015**, *16*, 876. [[CrossRef](#)] [[PubMed](#)]
44. Gardiner, E.J.; Cairns, M.J.; Liu, B.; Beveridge, N.J.; Carr, V.; Kelly, B.; Scott, R.J.; Tooney, P.A. Gene expression analysis reveals schizophrenia-associated dysregulation of immune pathways in peripheral blood mononuclear cells. *J. Psychiatr. Res.* **2013**, *47*, 425–437. [[CrossRef](#)] [[PubMed](#)]
45. Pramparo, T.; Pierce, K.; Lombardo, M.V.; Carter Barnes, C.; Marinero, S.; Ahrens-Barbeau, C.; Murray, S.S.; Lopez, L.; Xu, R.; Courchesne, E. Prediction of Autism by Translation and Immune/Inflammation Coexpressed Genes in Toddlers from Pediatric Community Practices. *JAMA Psychiatry* **2015**, *72*, 386. [[CrossRef](#)]
46. Van Eijk, K.R.; De Jong, S.; Strengman, E.; Buizer-Voskamp, J.E.; Kahn, R.S.; Boks, M.P.; Horvath, S.; Ophoff, R.A. Identification of schizophrenia-associated loci by combining DNA methylation and gene expression data from whole blood. *Eur. J. Hum. Genet.* **2015**, *23*, 1106–1110. [[CrossRef](#)] [[PubMed](#)]
47. Leday, G.G.R.; Vértes, P.E.; Richardson, S.; Greene, J.R.; Regan, T.; Khan, S.; Henderson, R.; Freeman, T.C.; Pariante, C.M.; Harrison, N.A.; et al. Replicable and Coupled Changes in Innate and Adaptive Immune Gene Expression in Two Case-Control Studies of Blood Microarrays in Major Depressive Disorder. *Biol. Psychiatry* **2018**, *83*, 70–80. [[CrossRef](#)] [[PubMed](#)]
48. Shamir, R.; Klein, C.; Amar, D.; Vollstedt, E.J.; Bonin, M.; Usenovic, M.; Wong, Y.C.; Maver, A.; Poths, S.; Safer, H.; et al. Analysis of blood-based gene expression in idiopathic Parkinson disease. *Neurology* **2017**, *89*, 1676–1683. [[CrossRef](#)] [[PubMed](#)]
49. Sood, S.; Gallagher, I.J.; Lunnon, K.; Rullman, E.; Keohane, A.; Crossland, H.; Phillips, B.E.; Cederholm, T.; Jensen, T.; van Loon, L.J.C.; et al. A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biol.* **2015**, *16*, 185. [[CrossRef](#)]
50. Clelland, C.L.; Read, L.L.; Panek, L.J.; Nadrich, R.H.; Bancroft, C.; Clelland, J.D. Utilization of Never-Medicating Bipolar Disorder Patients towards Development and Validation of a Peripheral Biomarker Profile. *PLoS ONE* **2013**, *8*, e69082. [[CrossRef](#)]
51. Gazestani, V.H.; Pramparo, T.; Nalabolu, S.; Kellman, B.P.; Murray, S.; Lopez, L.; Pierce, K.; Courchesne, E.; Lewis, N.E. A perturbed gene network containing PI3K–AKT, RAS–ERK and WNT– β -catenin pathways in leukocytes is linked to ASD genetics and symptom severity. *Nat. Neurosci.* **2019**, *22*, 1624–1634. [[CrossRef](#)] [[PubMed](#)]
52. Osborne, L.; Clive, M.; Kimmel, M.; Gispen, F.; Guintivano, J.; Brown, T.; Cox, O.; Judy, J.; Meilman, S.; Braier, A.; et al. Replication of epigenetic postpartum depression biomarkers and variation with hormone levels. *Neuropsychopharmacology* **2016**, *41*, 1648–1658. [[CrossRef](#)]
53. Sabunciyani, S.; Maher, B.; Bahn, S.; Dickerson, F.; Yolken, R.H. Association of DNA methylation with acute mania and inflammatory markers. *PLoS ONE* **2015**, *10*, e0132001. [[CrossRef](#)] [[PubMed](#)]
54. Vallerga, C.L.; Zhang, F.; Fowdar, J.; McRae, A.F.; Qi, T.; Nabais, M.F.; Zhang, Q.; Kassam, I.; Henders, A.K.; Wallace, L.; et al. Analysis of DNA methylation associates the cystine–glutamate antiporter SLC7A11 with risk of Parkinson’s disease. *Nat. Commun.* **2020**, *11*, 1238. [[CrossRef](#)] [[PubMed](#)]
55. Ratanatharathorn, A.; Boks, M.P.; Maihofer, A.X.; Aiello, A.E.; Amstadter, A.B.; Ashley-Koch, A.E.; Baker, D.G.; Beckham, J.C.; Bromet, E.; Dennis, M.; et al. Epigenome-wide association of PTSD from heterogeneous cohorts with a common multi-site analysis pipeline. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **2017**, *174*, 619–630. [[CrossRef](#)]

56. Chuang, Y.-H.; Paul, K.C.; Bronstein, J.M.; Bordelon, Y.; Horvath, S.; Ritz, B. Parkinson's disease is associated with DNA methylation levels in human blood and saliva. *Genome Med.* **2017**, *9*, 76. [[CrossRef](#)]
57. Lohoff, F.W.; Roy, A.; Jung, J.; Longley, M.; Rosoff, D.B.; Luo, A.; O'Connell, E.; Sorcher, J.L.; Sun, H.; Schwandt, M.; et al. Epigenome-wide association study and multi-tissue replication of individuals with alcohol use disorder: Evidence for abnormal glucocorticoid signaling pathway gene regulation. *Mol. Psychiatry* **2020**, *26*, 2224–2237. [[CrossRef](#)] [[PubMed](#)]
58. Booij, L.; Casey, K.F.; Antunes, J.M.; Szyf, M.; Joobor, R.; Israël, M.; Steiger, H. DNA methylation in individuals with anorexia nervosa and in matched normal-eater controls: A genome-wide study. *Int. J. Eat. Disord.* **2015**, *48*, 874–882. [[CrossRef](#)]
59. Hannon, E.; Dempster, E.; Viana, J.; Burrage, J.; Smith, A.R.; Macdonald, R.; St Clair, D.; Mustard, C.; Breen, G.; Therman, S.; et al. An integrated genetic-epigenetic analysis of schizophrenia: Evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol.* **2016**, *17*, 176. [[CrossRef](#)]
60. Glatting, G.; Kletting, P.; Reske, S.N.; Hohl, K.; Ring, C. Choosing the optimal fit function: Comparison of the Akaike information criterion and the F-test. *Med. Phys.* **2007**, *34*, 4285–4292. [[CrossRef](#)] [[PubMed](#)]
61. Zhao, N.; Liu, C.C.; Qiao, W.; Bu, G. Apolipoprotein E, Receptors, and Modulation of Alzheimer's Disease. *Biol. Psychiatry* **2018**, *83*, 347–357. [[CrossRef](#)] [[PubMed](#)]
62. Uhlen, M.; Karlsson, M.J.; Zhong, W.; Tebani, A.; Pou, C.; Mikes, J.; Lakshmikanth, T.; Forsström, B.; Edfors, F.; Odeberg, J.; et al. A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **2019**, *366*, eaax9198. [[CrossRef](#)] [[PubMed](#)]
63. Dickens, A.M.; Sen, P.; Kempton, M.J.; Barrantes-Vidal, N.; Iyegbe, C.; Nordentoft, M.; Pollak, T.; Riecher-Rössler, A.; Ruhrmann, S.; Sachs, G.; et al. Dysregulated Lipid Metabolism Precedes Onset of Psychosis. *Biol. Psychiatry* **2020**, *89*, 288–297. [[CrossRef](#)] [[PubMed](#)]
64. Vogelzangs, N.; de Jonge, P.; Smit, J.H.; Bahn, S.; Penninx, B.W. Cytokine production capacity in depression and anxiety. *Transl. Psychiatry* **2016**, *6*, e825. [[CrossRef](#)] [[PubMed](#)]
65. Hartwig, F.P.; Borges, M.C.; Horta, B.L.; Bowden, J.; Davey Smith, G. Inflammatory Biomarkers and Risk of Schizophrenia. *JAMA Psychiatry* **2017**, *74*, 1226. [[CrossRef](#)] [[PubMed](#)]