


Article

Integrated Analysis of Methylomic and Transcriptomic Data to Identify Potential Diagnostic Biomarkers for Major Depressive Disorder

Yinping Xie ¹, Ling Xiao ², Lijuan Chen ³ , Yage Zheng ¹, Caixia Zhang ⁴ and Gaohua Wang ^{1,2,*}

¹ Department of Psychiatry, Renmin Hospital of Wuhan University, Wuhan 430060, China; yinpingxie@whu.edu.cn (Y.X.); yagezheng2012@sina.com (Y.Z.)

² Institute of Neuropsychiatry, Renmin Hospital, Wuhan University, Wuhan 430060, China; anemous7851@hotmail.com

³ College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China; chen_lijuanrabbit@126.com

⁴ Department of Cardiology, Renmin Hospital of Wuhan University, Wuhan 430060, China; 13835198286@163.com

* Correspondence: wgh6402@whu.edu.cn; Tel.: +86-27-88041911

Abstract: Major depressive disorder (MDD) is a mental illness with high incidence and complex etiology, that poses a serious threat to human health and increases the socioeconomic burden. Currently, high-accuracy biomarkers for MDD diagnosis are urgently needed. This paper aims to identify novel blood-based diagnostic biomarkers for MDD. Whole blood DNA methylation data and gene expression data from the Gene Expression Omnibus database are downloaded. Then, differentially expressed/methylated genes (DEGs/DMGs) are identified. In addition, we made a systematic analysis of the DNA methylation on 5'-C-phosphate-G-3' (CpGs) in all of the gene regions, as well as different gene regions, and then we defined a "dominant" region. Subsequently, integrated analysis is employed to identify the robust MDD-related blood biomarkers. Finally, a gene expression classifier and a methylation classifier are constructed using the random forest algorithm and the leave-one-out cross-validation method. Our results demonstrate that DEGs are mainly involved in the inflammatory response-associated pathways, while DMGs are primarily concentrated in the neurodevelopment- and neuroplasticity-associated pathways. Our integrated analysis identified 46 hypo-methylated and up-regulated (hypo-up) genes and 71 hyper-methylated and down-regulated (hyper-down) genes. One gene expression classifier and two DNA methylation classifiers, based on the CpGs in all of the regions or in the dominant regions are constructed. The gene expression classifier possessed the best predictive ability, followed by the DNA methylation classifiers, based on the CpGs in both the dominant regions and all of the regions. In summary, the integrated analysis of DNA methylation and gene expression has identified 46 hypo-up genes and 71 hyper-down genes, which could be used as diagnostic biomarkers for MDD.

Keywords: major depressive disorder; diagnostic biomarkers; DNA methylation; mRNA expression; random forest algorithm; leave-one-out cross-validation



Citation: Xie, Y.; Xiao, L.; Chen, L.; Zheng, Y.; Zhang, C.; Wang, G. Integrated Analysis of Methylomic and Transcriptomic Data to Identify Potential Diagnostic Biomarkers for Major Depressive Disorder. *Genes* **2021**, *12*, 178. <https://doi.org/10.3390/genes12020178>

Received: 7 November 2020

Accepted: 26 January 2021

Published: 27 January 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

Major depressive disorder (MDD) is one of the most common mental disorders around the world, and as estimated by the World Health Organization, approximately 350 million people of all ages worldwide suffer from depression [1]. The symptoms of depression include pervasive and persistent low mood, lack of motivation, and loss of interest in social interactions, which are important public health problems contributing to severe morbidity and mortality [2–4]. At present, the most common classical method for diagnosing depression is scale assessment, and imagological diagnosis may provide effective information in MDD classification [5,6]. Recently, as a relatively low-invasive and accessible method, peripheral blood (PB) examination

has become an important complement to the classic diagnostic method mentioned above, improving the accuracy of an MDD diagnosis [7]. For example, microRNAs, exosomes, or a certain protein (e.g., C-reactive protein) in PB samples have been used as powerful tools to distinguish MDD patients from healthy controls [8,9]. To date, high-accuracy biomarkers for MDD diagnosis and prognosis are still lacking. Therefore, it is of great significance to investigate the molecular mechanism of MDD, aiming to identify precise targets and necessary biomarkers for the diagnosis of MDD.

MDD is a complex and heterogeneous disease strongly associated with genetic and environmental factors [10]. Each genetic or environmental factor alone cannot sufficiently explain MDD [11,12]. This then motivates people in the field of MDD research to select epigenetic mechanisms as prime candidates for mediating the genetic and environmental interactions in several brain regions [13]. Furthermore, DNA methylation is one of the major epigenetic modifications and plays an important role in the etiology of complex diseases [14]. Thus far, most DNA methylation studies have used candidate gene approaches and have been predominantly focused on gene promoter regions [15–17]. Accompanied by the rise of DNA methylation chip arrays and whole-genome bisulfite sequencing technology, several attempts have been made to decipher the relationship between DNA methylation and depression [18,19]. Although many genome-wide studies have indicated that DNA methylation is associated with depression, both positive and negative associations have been reported, and conflicting results are often observed [20,21].

DNA methylation is a very complicated phenomenon. It can occur in different regions such as in transcriptional start sites (TSS), gene bodies, and beyond. Except for the most studied promoter region, the function of other regions has been a largely underexplored domain. Perhaps the average methylation level of a specific gene is not a good reflection of methylation and disease, and a certain region or a CpG site may be better. In this study, we made a systematic analysis of DNA methylation on CpGs in all gene regions, as well as different gene regions (i.e., TSS1500, TSS200, 5' untranslated region (UTR), first exon (Exon1st), gene body, and 3'UTR), and then defined a “dominant” region, making the DNA methylation research in MDD more elaborate.

Both DNA methylation and gene expression are associated with depression, and conventional wisdom holds that DNA methylation has a negative regulatory effect on gene expression. Their combined analysis gives us a more in-depth understanding of MDD. Hence, we aimed to screen for genes associated with methylation alterations, as well as gene expression changes, through integrated analysis to provide more accurate diagnosis biomarkers for MDD. In this study, integrated analysis of DNA methylation and gene expression data identified 46 hypo-methylated and up-regulated (hypo-up) genes and 71 hyper-methylated and down-regulated (hyper-down) genes, and the random forest (RF) algorithm and leave-one-out (LOO) cross-validation method indicated that they could be used as diagnostic biomarkers for MDD.

2. Materials and Methods

2.1. Data Collection

The Gene Expression Omnibus (GEO) database is the largest and most comprehensive public gene expression data resource archiving and sorting high-throughput gene expression and genomics data. Our goal was to identify diagnostic biomarkers through a relatively accessible and low-invasive method. Therefore, we used the keywords “MDD” or “major depressive disorder” and “blood sample” for retrieval from GEO. After a careful review, the DNA methylation dataset GSE113725, including blood samples of 100 MDD patients and 50 healthy controls, was downloaded from the GEO database (GPL13534 Illumina HumanMethylation450 BeadChip) [22]. In addition, a gene expression profile, GSE98793, deposited by Leday et al., was selected based on the GPL570 HG-U133_Plus_2 platform, containing blood samples of 128 MDD patients and 64 healthy controls [23]. The demographic and clinical features for GSE113725 and GSE98793 are listed in Supplementary Tables S1 and S2. The workflow of this study is shown in Figure 1.

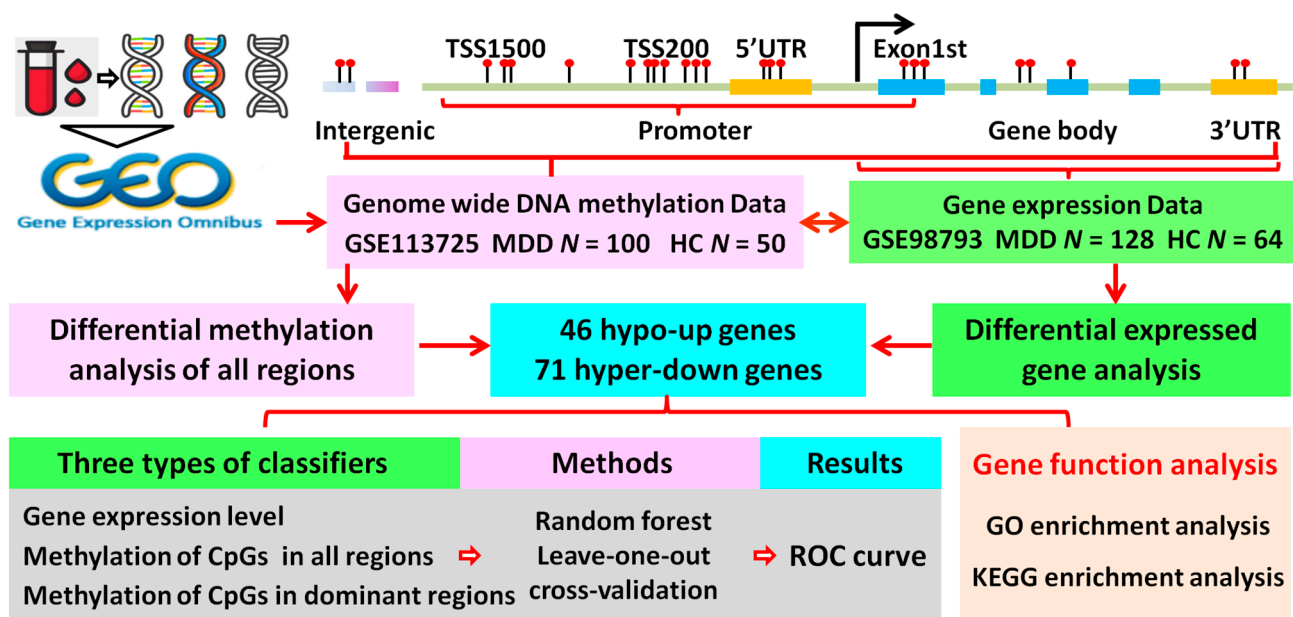


Figure 1. Flowchart of the analysis process. CpGs, 5'-C-phosphate-G-3'; GO, Gene Ontology; HC, healthy control; hyper-down, hyper-methylated and down-regulated; hypo-up, hypo-methylated and up-regulated; KEGG, Kyoto Encyclopedia of Genes and Genomes; MDD, major depressive disorder; ROC, receiver operating characteristic; TSS, transcriptional start site; UTR, untranslated region. *N*, number of MDD or HC.

2.2. Screening of Differentially Expressed Genes (DEGs)

The “Affy” package in the R software (version 3.5.2) was adopted to process the raw data in CEL format. After eliminating batch differences and performing data background correction, normalization, and summarization, a robust multiarray average was created for further analysis. The “limma” package was applied to assess the differential expression between MDD patients and healthy controls [24]. Benjamini and Hochberg’s (BH) method was used to control the false discovery rate across all genes. The threshold for identifying of DEGs was set at a BH-adjusted *p*-value of <0.05 and a $|\text{Log}_2 \text{fold-change}| > 0.2$.

2.3. Differential Methylation Analysis

Illumina Infinium Human Methylation450 BeadChip array covering 99% of the genes’ annotated promoter (TSS1500, TSS200), 5’UTR, Exon1st, gene body, and 3’UTR in the RefGENE database is one of the most classic DNA methylation detection techniques. TSS200/TSS1500 stands for 0-200/200-1500 bases upstream of the TSS, while gene body refers to the region between the initiation codon and stop codon. The analysis process is as follows:

- (1) **Methylation level of each CpG.** The methylation level of each CpG can be calculated by the equation $\beta = M / (M + U + a)$, where $M > 0$, $U > 0$, and $a \geq 0$. *M* and *U* denote the number of methylated and unmethylated probes, respectively. Since *M* and *U* are small, “*a*” is set to 100 to stabilize the β -value [25].
- (2) **Methylation level of different regions.** In this study, we employed the “ChAMP” package (version 2.18.3) to measure the methylation level of the different regions (TSS1500, TSS200, 5’UTR, Exon1st, gene body, and 3’UTR) for each individual gene using the average β -value of the CpGs in the corresponding regions.
- (3) **Methylation level of an individual gene.** We also measured the methylation level of an individual gene using the average β -value of the CpGs in all regions.
- (4) **Identification of differentially methylated genes (DMGs).** To measure the methylation difference between MDD patients and healthy controls, a linear model was built. Ten quantiles of the delta beta value of all genes and all intergenic CpG sites were

calculated, and the DMGs were defined as a $\Delta\beta$ value of $<1/10$ quantile or $>9/10$ quantile and a BH-adjusted p -value of <0.05 .

2.4. Identification of the *Dominant Hypo/Hyper-Methylated Regions*

In this study, we defined the dominant hypo/hyper-methylated regions (hereafter, dominant regions). A dominant region refers to the smallest delta beta value between the MDD patients and healthy controls. It is worth noting that there may be more than one dominant region in an individual gene. Herein, if the difference between the delta beta value of another region and the smallest delta beta value was smaller than 0.005, the region was regarded as one of the dominant regions [26].

2.5. GO and KEGG Enrichment Analyses

To better understand the biological functions of the DEGs and DMGs, GO enrichment analysis was performed to provide structured annotations on three subontologies: Biological process (BP), molecular function (MF), and cellular component (CC). KEGG pathway enrichment analysis of the DEGs and DMGs was also implemented. Both enrichment analyses employed the R package “clusterProfiler”. A BH-adjusted p -value of <0.05 was set as the cut-off criterion.

2.6. Classifier Construction and LOO Validation

RF machine learning is a nonlinear classifier that trains a large number of decision trees and uses the class predicted the most from these trees as the final prediction, which has been widely used in bioinformatics analysis, such as in in vivo transcription factor-binding prediction [27], and enhancer identification [28]. For this study, RF, implemented by the R package “randomForest,” was employed to build prediction models to distinguish MDD patients from healthy controls on 46 hypo-up genes and 71 hyper-down genes. Three types of classifiers were trained based on the log2 transformed gene expression level data, the average β -value of the CpGs in all of the regions, and the average β -value of the CpGs in the dominant regions.

2.7. LOO Cross-Validation

LOO is a cross-validation method that removes only one sample from the training set, and each learning set is created by taking all of the samples except one (test set left out). We employed the LOO cross-validation method to monitor the performance of the classifiers using the R package “caret.” The discriminative ability of each classifier was measured by receiver operating characteristic (ROC) curves, and the area under the ROC curve (AUROC) was calculated using the R package “ROCR.”

3. Results

3.1. Identification of the DEGs in MDD

To identify the DEGs in the MDD patients and healthy controls, we selected the most frequently used dataset, namely, GSE98793, containing blood samples of 128 MDD patients and 64 healthy controls. By employing the linear modeling approach, a total of 1506 DEGs were identified, 713 of which were up-regulated and the other 793 down-regulated (Figure 2A and Table S3). The top 50 genes with significant differences in up- and down-regulated genes were selected to construct a heat map to show the changes in the DEG expression (Figure 2B).

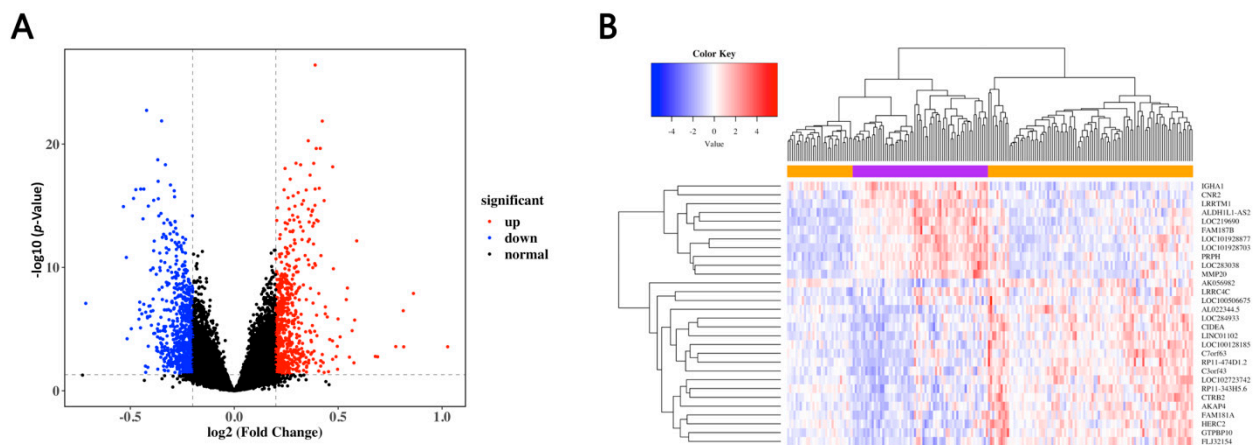


Figure 2. Identification of the DEGs in GES98793. **(A)** Volcano plots of differentially expressed genes (DEGs) in the MDD patients and healthy controls. The red and blue dots represent the up-regulated and down-regulated genes, respectively, while the black dots refer to the non-DEGs. **(B)** Heat map of the top 50 DEGs in the MDD patients and healthy controls.

3.2. GO and KEGG Enrichment Analysis of the DEGs

To explore the biological relevance of the DEGs, we performed GO and KEGG enrichment analysis and found that the DEGs were associated with the following: (1) BP terms: Neutrophil-mediated immunity, neutrophil activation involved in immune response, antimicrobial humoral response, etc.; (2) CC terms: Specific granule lumen, secretory granule lumen, cytoplasmic vesicle lumen, etc.; (3) MF terms: Serine-type endopeptidase activity, serine-type peptidase activity, hydrolase activity, acting on acid phosphorus–nitrogen bonds, etc.; (4) KEGG pathways: Hematopoietic cell lineage, asthma, complement, and coagulation cascades, etc. (Figure 3A–D). These results tie in well with previous studies, wherein inflammation was shown to trigger depression [29–31]. Proinflammatory cytokines, including IL-1, IL-6, and TNF- α , exhibited higher circulating levels in MDD patients than in non-depressed individuals [32]. The results of our analysis may provide new inflammation-associated diagnostic biomarkers for depression.

3.3. Identification and GO/KEGG Enrichment Analysis of the DMGs

We identified the significant DMGs based on the linear modeling approach and the delta β value of the CpGs of all of the regions. A total of 8313 DMGs were identified, including 4636 hyper-methylated genes and 3677 hypo-methylated genes (Supplementary Table S4). We also performed GO and KEGG enrichment analysis of all of the DMGs, and the results showed that the DMGs were associated with following: (1) BP terms: Axonogenesis, neuron projection guidance, axon guidance, etc.; (2) CC terms: Cell leading edge, synaptic membrane, cell–substrate junction, etc.; (3) MF terms: Actin binding, DNA-binding transcription activator activity, DNA-binding transcription activator activity, RNA polymerase II-specific, etc.; (4) KEGG pathways: Axon guidance, MAPK signaling pathway, Rap1 signaling pathway, etc. (Figure 4A–D). The results highlighted that the circulating DNA methylation probably participates in neurodevelopment and neuroplasticity, which are significantly associated with MDD.

3.4. Integrated Analysis of the Gene Expression and DNA Methylation

It is generally considered that there is a negative regulatory relationship between DNA methylation and gene expression. Therefore, we obtained the overlap of hypo-methylated and up-regulated genes, as well as hyper-methylated and down-regulated genes. As a result, 46 hypo-up genes and 71 hyper-down genes were identified (Figure 5A,B). All of the gene symbols are listed in Table 1. The heat map showed the changes in the hypo-up and hyper-down genes between the MDD patients and healthy controls (Figure 5C,D).

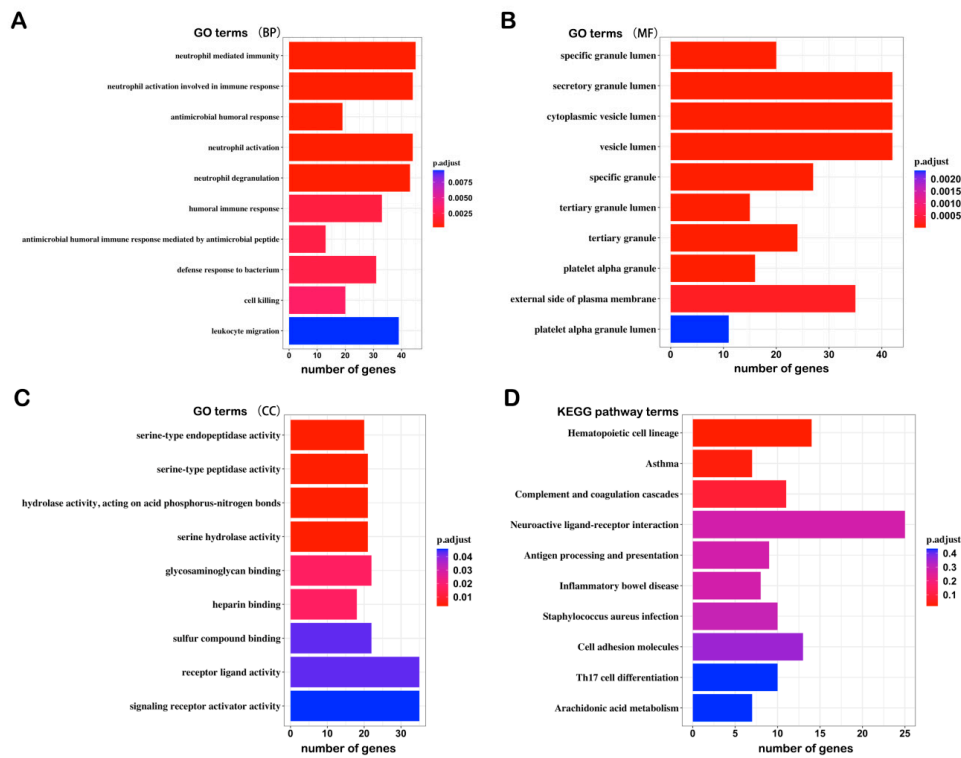


Figure 3. Top 10 GO and KEGG terms from the pathway enrichment analysis of the DEGs: (A) Biological process (BP); (B) molecular function (MF); (C) cellular component (CC); (D) KEGG pathway.

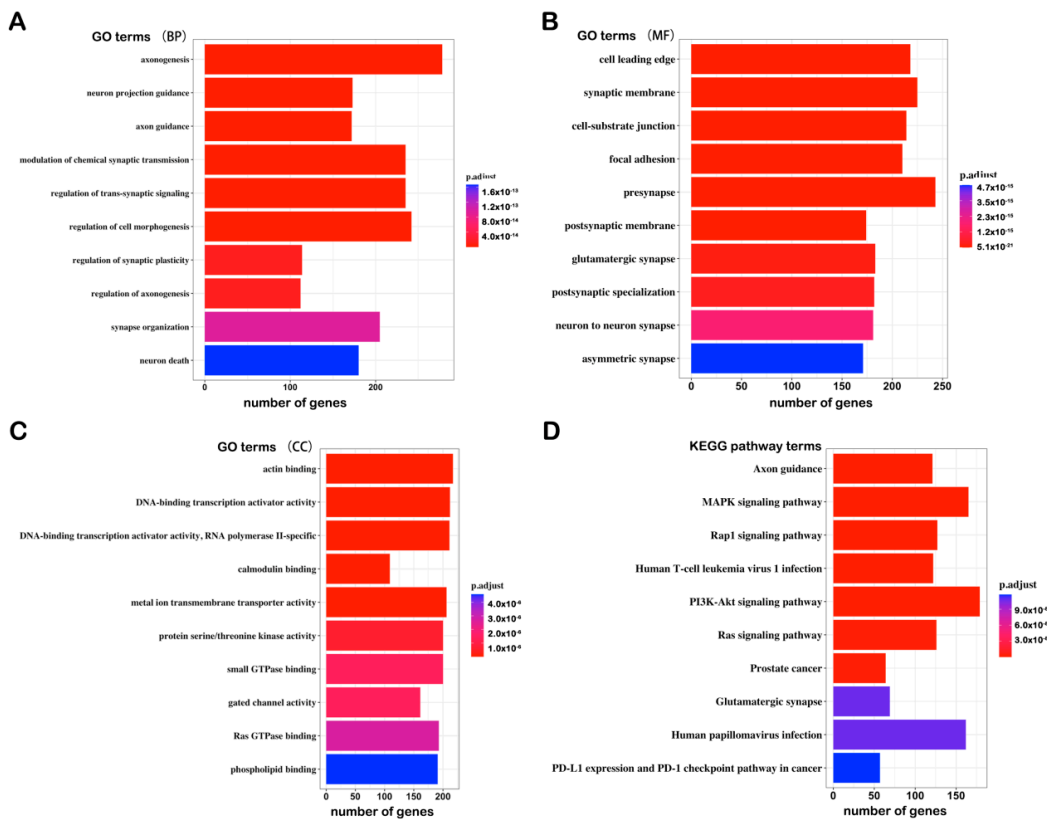


Figure 4. Top 10 GO and KEGG terms of the analysis of the differentially methylated genes (DMGs): (A) Biological process; (B) molecular function; (C) cellular component; (D) KEGG pathway.

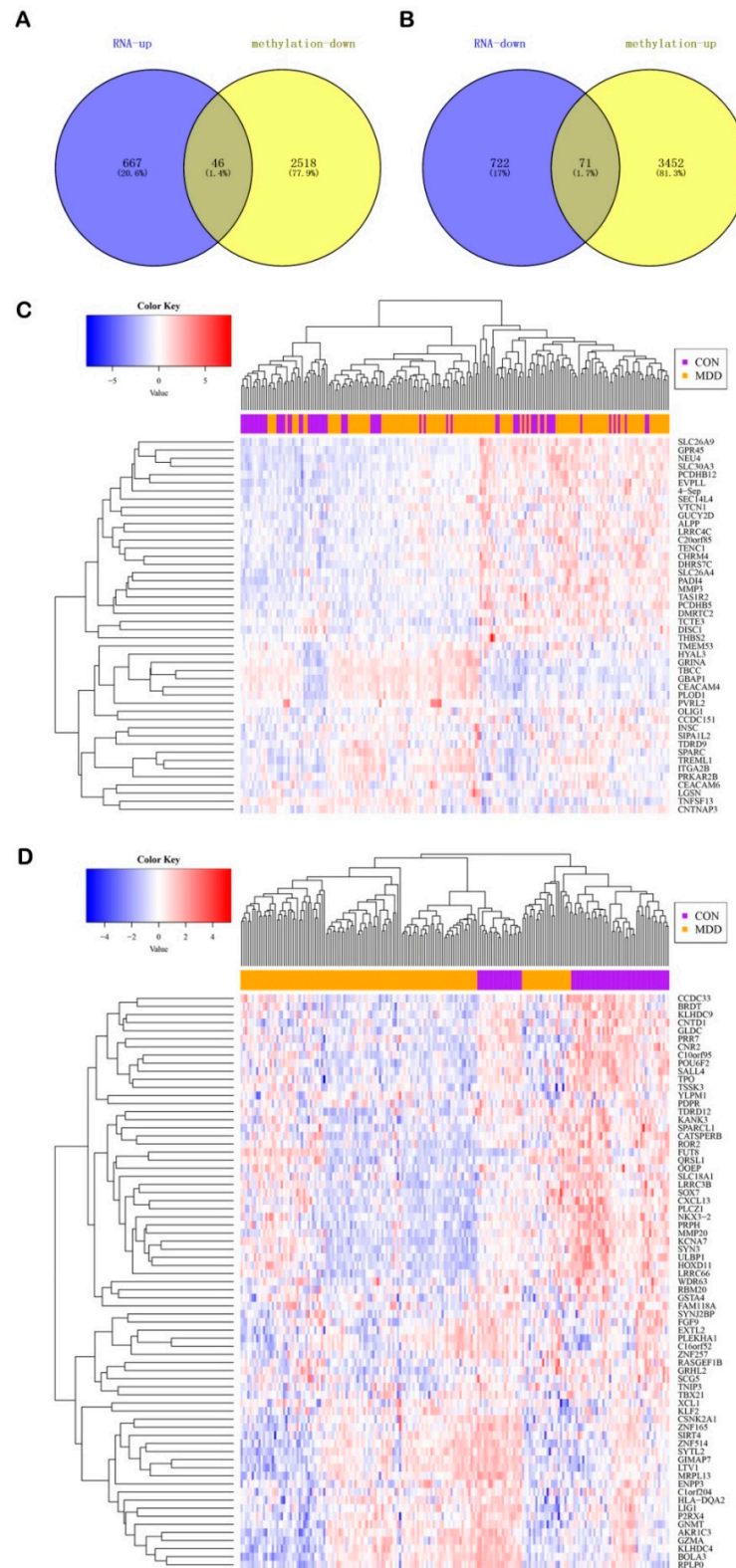


Figure 5. Identification of the hypo-up genes and hyper-down genes. Venn diagram of (A) the hypo-up genes and (B) the hyper-down genes. Heatmap of (C) the hypo-up genes and (D) the hyper-down genes between the MDD patients and healthy controls.

Table 1. Gene symbol of 46 hypo-up genes and 71 hyper-down genes.

Direction	Gene Symbol
hypo-up	<i>SEPT4, ALPP, C20orf85, CCDC151, CEACAM4, CEACAM6, CHRM4, CNTNAP3, DHRS7C, DISC1, DMRTC2, EVPLL, GBAP1, GPR45, GRINA, GUCY2D, HYAL3, INSCI, TGA2B, LGSN, LRRC4C, MMP3, NEU4, OLIG1, PADI4, PCDHB12, PCDHB5, PLOD1, PRKAR2B, PVRL2, SEC14L4, SIPA1L2, SLC26A4, SLC26A9, SLC30A3, SPARC, TAS1R2, TBCC, TCTE3, TDRD9, TENC1, THBS2, TMEM53, TNFSF13, TREML1, VTCN1</i>
hyper-down	<i>AKR1C3, BOLA3, BRDT, C10orf95, C16orf52, C1orf204, CATSPERB, CCDC33, CNR2, CNTD1, CSNK2A1CXCL13, ENPP3, EXTL2, FAM118A, FGF9, FUT8, GIMAP7, GLDC, GNMT, GRHL2, GSTA4, GZMA, HLA-DQA2, HOXD11, KANK3, KCNA7, KLF2, KLHDC4, KLHDC9, LIG1, LRRC3B, LRRC66, LTV1, MMP20, MRPL13, NKX3-2, OOE, P2RX4, PDPR, PLCZ1, PLEKHA1, POU6F2, PRPH, PRR7, QRSL1, RASGEF1B, RBM20, ROR2, RPLP0, SALL4, SCG5, SIRT4, SLC18A1, SOX7, SPARCL1, SYN3, SYNJ2BP, SYTL2, TBX21, TDRD12, TNIP3, TPO, TSSK3, ULBP1, WDR63, XCL1, YLPM1, ZNF165, ZNF257, ZNF514</i>

The 46 hypo-up genes were involved in such pathways (Table S5) as PI3K–Akt signaling pathway, the IL-17 signaling pathway, axon guidance, and neuroactive ligand–receptor interaction. The 71 hyper-down genes were involved in pathways (Table S6), such as the NF-kappa B signaling pathway, the MAPK signaling pathway, neuroactive ligand–receptor interaction, and the synaptic vesicle cycle. It is worth noting that most of the 46 hypo-up genes and 71 hyper-down genes were associated with both the inflammatory response and neuroplasticity. We also found that a proportion of the 71 hyper-down genes were engaged in nucleotide excision repair, DNA replication, ribosome, and ribosome biogenesis in the eukaryotes pathways. The results reveal that depression should be accompanied by changes in the metabolism of biological macromolecules, such as DNA and proteins.

3.5. Identification of the DMGs Based on CpGs in the Different Regions

In a certain gene, DNA methylation occurs in different regions, including TSS1500, TSS200, 1stExon, gene body, 5'UTR, and 3'UTR, as well as other regions, the function of which remains unclear. Genomic annotation of the methylation based on the CpGs in different regions revealed a biased genomic distribution. As shown in Figure 6A, the maximum number of DMGs is in the gene body region (>3000), followed by the TSS1500 region (>2000), while the 1stExon region had the minimum (<500). The DNA methylation distribution tendency of the overlap of the DEGs and DMGs was identical in that the gene body region had more than 100 DMGs ranking first, with the TSS1500 region coming second with more than 80, and the 1stExon region being last with at least 30 (Figure 6B). In terms of individual regions, the gene body had the highest percentage of DMGs. Unfortunately, little research on the DNA methylation in this region has been reported.

To get a closer look at the distribution of each region, we divided all of the overlaps of DMGs and DEGs into four groups: Hyper-up group, hyper-down group, hypo-up group, and hypo-down group. The methylation distribution of the four groups in each region is listed in Table 2. Figure 6C shows that the distribution of the CpGs of the hyper-down and hypo-up genes in the different regions is quite similar. The gene body and TSS1500 region remained in the top two, while the least in the hyper-down group was in 1stExon and in the hypo-up group was TSS200. The hyper-up group had the largest number of DMGs, and the DNA methylation distribution of the hypo-down group in the six regions was similar to that of the hyper-up group. We also found that the hyper-up group had the largest number in the gene body region (>60). It has been reported that the methylation of the gene body may have a positive impact on gene expression [33]. The relationship between gene body methylation and gene expression remains to be further elucidated. Our results suggest that region-specific methylation may play a potential role in the diagnosis of depression.

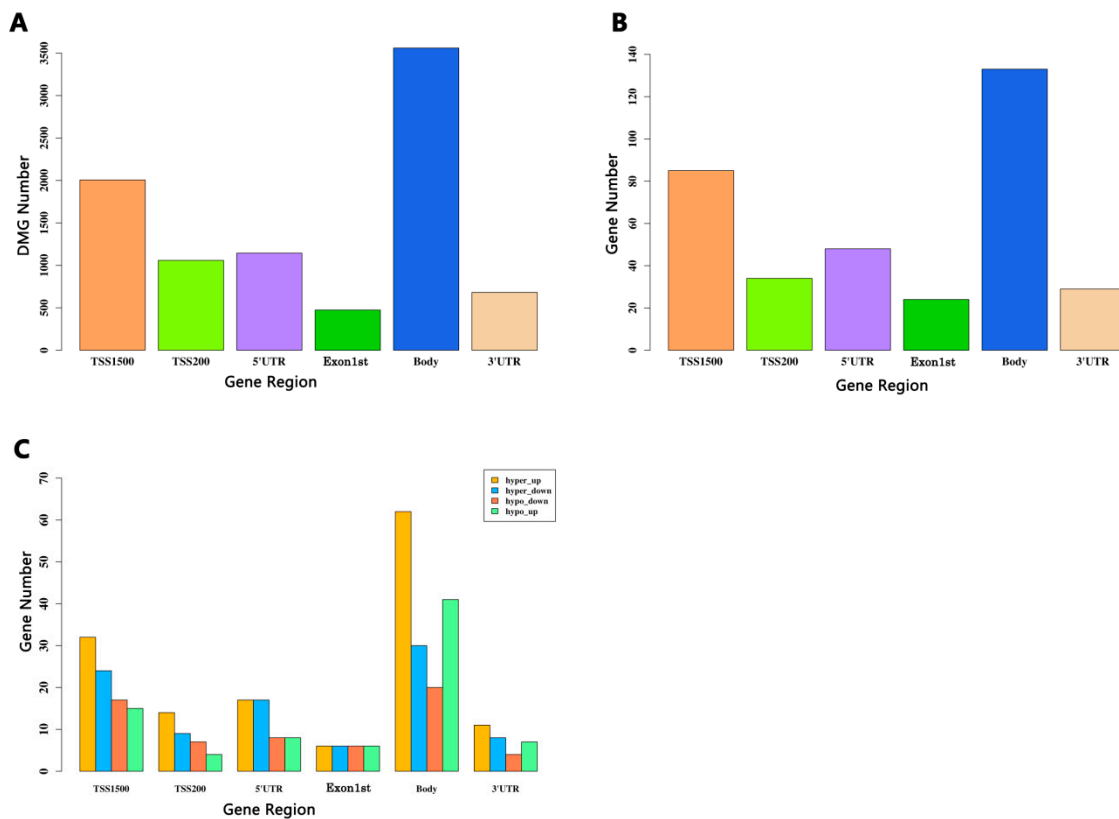


Figure 6. Integrated analysis of the DEGs and DMGs based on the CpGs in different regions. (A) Barplot for the CpG distribution of the DMGs. (B) Barplot for the overlapped genes between the DEGs and the different regions of the DMGs. The y-axis stands for the overlapped gene numbers. The x-axis represents different gene regions: TSS1500, TSS200, 5'UTR, Exon1st, body, and 3'UTR. (C) Barplot of the four groups that overlap in each region. Hyper-up represents the hyper-methylated and up-regulated genes. The y-axis is the number of genes. The x-axis represents different gene regions: TSS1500, TSS200, 5'UTR, Exon1st, body, and 3'UTR.

Table 2. Gene numbers of four groups in TSS1500, TSS200, 5'UTR, Exon1st, Body, 3'UTR region.

Group	TSS1500	TSS200	5'UTR	Exon1st	Body	3'UTR
hypo-up	15	4	8	6	41	7
hypo-down	17	7	8	6	20	4
hyper-up	32	14	17	6	62	11
hyper-down	24	9	17	6	30	8

3.6. Classifier Construction and ROC Curve

We employed the RF algorithm and the LOO cross-validation method to construct classifiers based on the gene expression and methylation data of the 46 hypo-up genes and 71 hyper-down genes to distinguish the MDD patients from the healthy controls. There were two types of methylation classifiers for the 46 hypo-up genes and 71 hyper-down genes based on the CpGs in all of the regions and the CpGs in the dominant regions only. The relevant information about the methylation of the CpGs in all regions, as well as the CpGs in the dominant regions, of the 46 hypo-up genes are shown in Supplementary Tables S7 and S8, while for the 71 hyper-down genes, the information is presented in Supplementary Tables S9 and S10.

3.6.1. The Importance Score of the 46 Hypo-Up Genes and the 71 Hyper-Down Genes in Each Classifier

The “importance” function of the “randomForest” package was used to calculate the average importance of each gene in the six classifiers, and their importance was ranked in descending order. Figure 7A–F shows the importance scores of the top 20 genes in each classifier, and all of the genes and importance scores are listed in Tables 3 and 4.

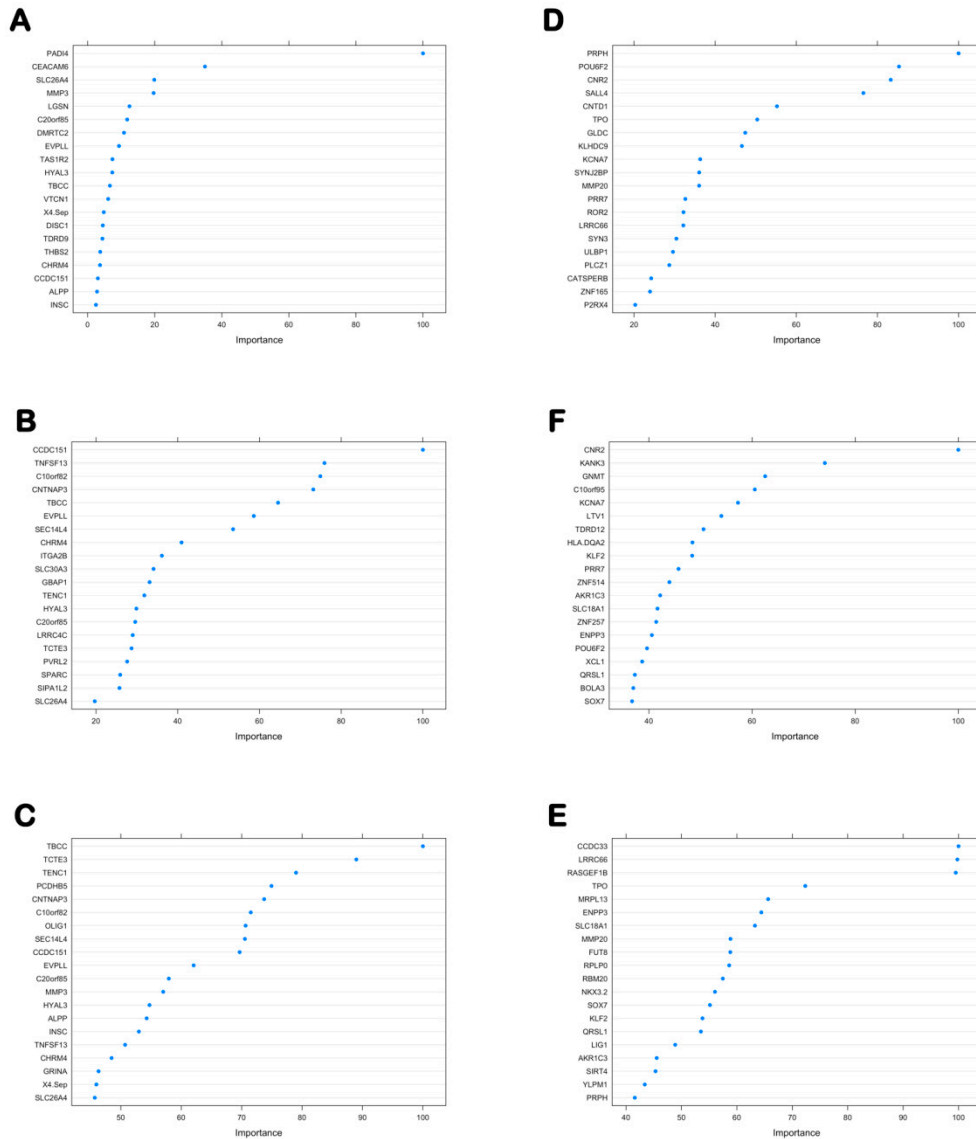


Figure 7. Top 20 genes and importance scores of the hypo-up genes and hyper-down genes in each classifier. Top 20 genes and importance scores of the hypo-up genes in (A) the gene expression classifier, (B) the gene methylation classifier based on CpGs in all of the regions, and (C) the gene methylation classifier based on the CpGs in the dominant regions. Top 20 genes and importance scores of the hyper-down genes in (D) the gene expression classifier, (E) the gene methylation classifier based on the CpGs in all of the regions, and (F) gene methylation classifier based on the CpGs in the dominant regions.

Table 3. Gene symbol and importance scores of 46 hypo-up genes in gene expression classifier and gene methylation classifiers.

Importance of 46 Hypo-Up Genes Based on Gene Expression Level		Importance of 46 Hypo-Up Genes Based on DNA Methylation Level of All Regions		Importance of 46 Hypo-Up Genes Based on DNA Methylation Level of Dominant Regions	
Gene Symbol	Importance	Gene Symbol	Importance	Gene Symbol	Importance
<i>PADI4</i>	100.00	<i>CCDC151</i>	100.00	<i>TBCC</i>	100.00
<i>CEACAM6</i>	34.98	<i>TNFSF13</i>	75.91	<i>TCTE3</i>	88.98
<i>SLC26A4</i>	19.87	<i>C10orf82</i>	74.91	<i>TENC1</i>	79.01
<i>MMP3</i>	19.66	<i>CNTNAP3</i>	73.18	<i>PCDHB5</i>	74.94
<i>LGSN</i>	12.47	<i>TBCC</i>	64.55	<i>CNTNAP3</i>	73.72
<i>C20orf85</i>	11.79	<i>EVPLL</i>	58.62	<i>C10orf82</i>	71.51
<i>DMRTC2</i>	10.84	<i>SEC14L4</i>	53.55	<i>OLIG1</i>	70.65
<i>EVPLL</i>	9.34	<i>CHRM4</i>	40.94	<i>SEC14L4</i>	70.53
<i>TAS1R2</i>	7.39	<i>ITGA2B</i>	36.11	<i>CCDC151</i>	69.67
<i>HYAL3</i>	7.34	<i>SLC30A3</i>	34.08	<i>EVPLL</i>	62.05
<i>TBCC</i>	6.63	<i>GBAP1</i>	33.11	<i>C20orf85</i>	57.94
<i>VTCN1</i>	6.12	<i>TENC1</i>	31.83	<i>MMP3</i>	57.02
<i>X4.SEP</i>	4.80	<i>HYAL3</i>	29.89	<i>HYAL3</i>	54.74
<i>DISC1</i>	4.52	<i>C20orf85</i>	29.60	<i>ALPP</i>	54.27
<i>TDRD9</i>	4.39	<i>LRRC4C</i>	28.99	<i>INSC</i>	52.99
<i>THBS2</i>	3.75	<i>TCTE3</i>	28.70	<i>TNFSF13</i>	50.71
<i>CHRM4</i>	3.68	<i>PVRL2</i>	27.63	<i>CHRM4</i>	48.47
<i>CCDC151</i>	3.04	<i>SPARC</i>	25.92	<i>GRINA</i>	46.32
<i>ALPP</i>	2.79	<i>SIPA1L2</i>	25.75	<i>X4.SEP</i>	45.95
<i>INSC</i>	2.47	<i>SLC26A4</i>	19.70	<i>SLC26A4</i>	45.69
<i>GPR45</i>	2.12	<i>OLIG1</i>	19.45	<i>PVRL2</i>	44.90
<i>PCDHB5</i>	2.09	<i>GPR45</i>	19.07	<i>TDRD9</i>	39.39
<i>ITGA2B</i>	2.01	<i>PCDHB12</i>	18.70	<i>DHRS7C</i>	38.25
<i>PLOD1</i>	2.00	<i>X4.Sep</i>	17.39	<i>PCDHB12</i>	38.25
<i>SLC30A3</i>	1.92	<i>TAS1R2</i>	17.30	<i>PADI4</i>	35.12
<i>PRKAR2B</i>	1.84	<i>SLC26A9</i>	17.11	<i>GBAP1</i>	34.98
<i>TMEM53</i>	1.69	<i>CEACAM4</i>	16.23	<i>CEACAM6</i>	33.94
<i>PCDHB12</i>	1.58	<i>CEACAM6</i>	16.17	<i>LGSN</i>	32.33
<i>GUCY2D</i>	1.52	<i>VTCN1</i>	16.08	<i>GUCY2D</i>	31.03
<i>TCTE3</i>	1.32	<i>MMP3</i>	15.67	<i>PLOD1</i>	30.28
<i>CEACAM4</i>	1.27	<i>TDRD9</i>	15.51	<i>PRKAR2B</i>	29.78
<i>TNFSF13</i>	1.26	<i>NEU4</i>	15.48	<i>NEU4</i>	29.59
<i>CNTNAP3</i>	1.26	<i>DMRTC2</i>	14.53	<i>THBS2</i>	28.19
<i>LRRC4C</i>	1.13	<i>PADI4</i>	13.87	<i>SIPA1L2</i>	26.04
<i>SEC14L4</i>	1.07	<i>GUCY2D</i>	12.33	<i>GPR45</i>	25.63
<i>DHRS7C</i>	1.07	<i>ALPP</i>	10.95	<i>VTCN1</i>	20.80
<i>SLC26A9</i>	1.04	<i>PLOD1</i>	10.59	<i>TMEM53</i>	18.90
<i>OLIG1</i>	0.98	<i>PCDHB5</i>	9.36	<i>DMRTC2</i>	18.00
<i>GRINA</i>	0.96	<i>THBS2</i>	9.34	<i>LRRC4C</i>	11.03
<i>PVRL2</i>	0.83	<i>PRKAR2B</i>	8.81	<i>SLC26A9</i>	10.08
<i>TENC1</i>	0.74	<i>DHRS7C</i>	7.63	<i>ITGA2B</i>	8.65
<i>NEU4</i>	0.59	<i>GRINA</i>	7.02	<i>SPARC</i>	8.55
<i>SIPA1L2</i>	0.53	<i>TMEM53</i>	6.53	<i>CEACAM4</i>	8.35
<i>SPARC</i>	0.43	<i>DISC1</i>	3.23	<i>SLC30A3</i>	7.98
<i>TREML1</i>	0.28	<i>LGSN</i>	2.52	<i>TAS1R2</i>	5.77
<i>GBAP1</i>	0.00	<i>INSC</i>	0.00	<i>DISC1</i>	0.00

Table 4. Gene symbol and importance scores of 71 hyper-down genes in gene expression classifier and gene methylation classifiers.

Importance of 71 Hyper-Down Genes Based on Gene Expression Level		Importance of 71 Hyper-Down Genes Based on DNA Methylation Level of All Regions		Importance of 71 Hyper-Down Genes Based on DNA Methylation Level of Dominant Regions	
Gene Symbol	Importance	Gene Symbol	Importance	Gene Symbol	Importance
<i>PRPH</i>	100.00	<i>CNR2</i>	100.00	<i>CCDC33</i>	100.00
<i>POU6F2</i>	85.32	<i>KANK3</i>	74.12	<i>LRRC66</i>	99.79
<i>CNR2</i>	83.29	<i>GNMT</i>	62.55	<i>RASGEF1B</i>	99.50
<i>SALL4</i>	76.54	<i>C10orf95</i>	60.54	<i>TPO</i>	72.35
<i>CNTD1</i>	55.23	<i>KCNA7</i>	57.28	<i>MRPL13</i>	65.64
<i>TPO</i>	50.37	<i>LTV1</i>	54.06	<i>ENPP3</i>	64.41
<i>GLDC</i>	47.41	<i>TDRD12</i>	50.59	<i>SLC18A1</i>	63.26
<i>KLHDC9</i>	46.57	<i>HLA.DQA2</i>	48.45	<i>MMP20</i>	58.84
<i>KCNA7</i>	36.31	<i>KLF2</i>	48.39	<i>FUT8</i>	58.82
<i>SYNJ2BP</i>	36.05	<i>PRR7</i>	45.75	<i>RPLP0</i>	58.62
<i>MMP20</i>	36.04	<i>ZNF514</i>	43.97	<i>RBM20</i>	57.46
<i>PRR7</i>	32.64	<i>AKR1C3</i>	42.20	<i>NKX3.2</i>	56.05
<i>ROR2</i>	32.18	<i>SLC18A1</i>	41.67	<i>SOX7</i>	55.14
<i>LRRC66</i>	32.14	<i>ZNF257</i>	41.42	<i>KLF2</i>	53.79
<i>SYN3</i>	30.43	<i>ENPP3</i>	40.57	<i>QRSL1</i>	53.50
<i>ULBP1</i>	29.59	<i>POU6F2</i>	39.64	<i>LIG1</i>	48.87
<i>PLCZ1</i>	28.68	<i>XCL1</i>	38.70	<i>AKR1C3</i>	45.54
<i>CATSPERB</i>	24.22	<i>QRSL1</i>	37.29	<i>SIRT4</i>	45.32
<i>ZNF165</i>	23.93	<i>BOLA3</i>	36.98	<i>YLPM1</i>	43.38
<i>P2RX4</i>	20.28	<i>SOX7</i>	36.74	<i>PRPH</i>	41.58
<i>GIMAP7</i>	19.69	<i>FGF9</i>	34.30	<i>PLEKHA1</i>	41.24
<i>NKX3.2</i>	19.57	<i>SYTL2</i>	34.13	<i>SYNJ2BP</i>	40.67
<i>CSNK2A1</i>	19.37	<i>TBX21</i>	33.91	<i>FGF9</i>	40.42
<i>LIG1</i>	18.73	<i>MRPL13</i>	33.58	<i>KLHDC9</i>	39.15
<i>C10orf95</i>	18.64	<i>TNIP3</i>	32.47	<i>SYTL2</i>	37.12
<i>HOXD11</i>	14.99	<i>C1orf204</i>	32.40	<i>CATSPERB</i>	35.66
<i>MRPL13</i>	14.29	<i>EXTL2</i>	32.37	<i>GSTA4</i>	31.31
<i>BRDT</i>	12.75	<i>GRHL2</i>	30.95	<i>GNMT</i>	30.86
<i>SYTL2</i>	12.32	<i>MMP20</i>	30.93	<i>KANK3</i>	30.20
<i>RBM20</i>	12.02	<i>RASGEF1B</i>	30.61	<i>ROR2</i>	28.88
<i>SPARCL1</i>	11.97	<i>CATSPERB</i>	29.19	<i>P2RX4</i>	28.65
<i>ZNF514</i>	11.36	<i>C16orf52</i>	27.04	<i>CXCL13</i>	27.97
<i>GRHL2</i>	11.27	<i>SIRT4</i>	25.34	<i>HLA.DQA2</i>	27.92
<i>SLC18A1</i>	11.15	<i>GLDC</i>	24.13	<i>BOLA3</i>	26.26
<i>GNMT</i>	10.75	<i>CXCL13</i>	24.12	<i>LTV1</i>	25.90
<i>LRRC3B</i>	10.62	<i>FUT8</i>	22.93	<i>GIMAP7</i>	25.73
<i>GZMA</i>	10.06	<i>SYNJ2BP</i>	21.97	<i>FAM118A</i>	25.50
<i>FAM118A</i>	9.29	<i>KLHDC9</i>	21.81	<i>GRHL2</i>	24.98
<i>TDRD12</i>	9.01	<i>KLHDC4</i>	21.53	<i>CSNK2A1</i>	24.97
<i>RPLP0</i>	8.89	<i>GIMAP7</i>	20.03	<i>HOXD11</i>	24.96
<i>CXCL13</i>	8.53	<i>TSSK3</i>	19.54	<i>BRDT</i>	24.18
<i>CCDC33</i>	8.53	<i>HOXD11</i>	18.13	<i>C10orf95</i>	23.17
<i>ZNF257</i>	8.03	<i>BRDT</i>	17.55	<i>C1orf204</i>	23.08
<i>C1orf204</i>	7.90	<i>PRPH</i>	16.54	<i>XCL1</i>	23.02
<i>BOLA3</i>	7.60	<i>CSNK2A1</i>	15.97	<i>TSSK3</i>	22.30
<i>TSSK3</i>	7.59	<i>TPO</i>	15.85	<i>CNR2</i>	21.92
<i>KANK3</i>	7.51	<i>NKX3.2</i>	14.99	<i>TNIP3</i>	21.08
<i>ENPP3</i>	6.89	<i>RBM20</i>	14.11	<i>WDR63</i>	20.39
<i>GSTA4</i>	6.80	<i>PDPR</i>	13.91	<i>OOEP</i>	19.69
<i>AKR1C3</i>	6.75	<i>LRRC66</i>	12.73	<i>KLHDC4</i>	19.50
<i>KLHDC4</i>	6.28	<i>SPARCL1</i>	12.28	<i>GLDC</i>	19.26
<i>EXTL2</i>	5.85	<i>CNTD1</i>	12.28	<i>SPARCL1</i>	19.17
<i>SIRT4</i>	5.50	<i>FAM118A</i>	12.22	<i>GZMA</i>	18.74

Table 4. Cont.

Importance of 71 Hyper-Down Genes Based on Gene Expression Level		Importance of 71 Hyper-Down Genes Based on DNA Methylation Level of All Regions		Importance of 71 Hyper-Down Genes Based on DNA Methylation Level of Dominant Regions	
Gene Symbol	Importance	Gene Symbol	Importance	Gene Symbol	Importance
<i>FGF9</i>	5.22	<i>CCDC33</i>	11.64	<i>PLCZ1</i>	18.74
<i>SCG5</i>	4.88	<i>YLPM1</i>	11.39	<i>SCG5</i>	17.37
<i>FUT8</i>	4.51	<i>RPLP0</i>	9.51	<i>POU6F2</i>	16.12
<i>OOEP</i>	4.45	<i>PLEKHA1</i>	9.49	<i>PDPR</i>	15.66
<i>TNIP3</i>	4.44	<i>LIG1</i>	9.40	<i>PRR7</i>	15.27
<i>PDPR</i>	4.43	<i>LRRC3B</i>	9.20	<i>EXTL2</i>	14.63
<i>PLEKHA1</i>	4.29	<i>ZNF165</i>	8.61	<i>ZNF165</i>	13.83
<i>SOX7</i>	3.97	<i>GSTA4</i>	7.98	<i>LRRC3B</i>	13.70
<i>C16orf52</i>	3.90	<i>SYN3</i>	7.58	<i>SYN3</i>	13.64
<i>RASGEF1B</i>	3.56	<i>SALL4</i>	7.46	<i>SALL4</i>	13.33
<i>WDR63</i>	3.54	<i>SCG5</i>	6.31	<i>ULBP1</i>	12.34
<i>TBX21</i>	2.92	<i>ULBP1</i>	6.12	<i>TDRD12</i>	12.16
<i>HLA.DQA2</i>	2.77	<i>OOEP</i>	5.27	<i>ZNF514</i>	8.22
<i>KLF2</i>	2.76	<i>WDR63</i>	4.40	<i>C16orf52</i>	7.70
<i>LTV1</i>	2.35	<i>PLCZ1</i>	3.34	<i>TBX21</i>	7.52
<i>QRSL1</i>	1.46	<i>GZMA</i>	2.84	<i>CNTD1</i>	7.11
<i>XCL1</i>	0.72	<i>ROR2</i>	0.52	<i>KCNA7</i>	5.55
<i>YLPM1</i>	0.00	<i>P2RX4</i>	0.00	<i>ZNF257</i>	0.00

3.6.2. Determine the Number of Genes with the Best Predictive Power in Each Classifier

To obtain the best classification predictive power of each classifier, we added the candidate genes into each classifier one-by-one in order of importance. Figure 8A–C shows that the top 25, top 12, and top 23 are the best predictors of the hypo-up gene expression classifier, the gene methylation classifier based on the CpGs in all of the regions, that based on the CpGs in the dominant regions, respectively. Figure 8D–F shows that the top 31, top 2, and top 18 are the best predictors of the hyper-down gene expression classifier, the gene methylation classifier based on the CpGs in all of the regions, and that based on the CpGs in the dominant regions, respectively.

3.6.3. The Predictive Ability of Each Classifier

The ROC curve shows that the gene expression classifier exhibited the best predictive ability ($AUC = 0.964$, $p = 1.1 \times 10^{-25}$) for the hypo-up genes. The predictive ability ($AUC = 0.712$, $p = 3.7 \times 10^{-5}$) of the methylation classifier based on the CpGs in the dominant regions performed slightly better than that of the classifier based on the CpGs of all of the regions ($AUC = 0.677$, $p = 5.3 \times 10^{-4}$) (Figure 9A–C). In regard to the hyper-down genes, the gene expression classifier still presented the best predictive ability ($AUC = 0.9993$, $p = 1.9 \times 10^{-29}$). The predictive ability of the two methylation classifiers was similar for the classifier based on the CpGs of all of the regions ($AUC = 0.712$, $p = 3.2 \times 10^{-5}$) and on the CpGs of the dominant regions ($AUC = 0.716$, $p = 2.2 \times 10^{-5}$) (Figure 9D–F). The results reveal that for both the hypo-up and hyper-down genes, the classifier based on the gene expression data possessed the best predictive ability ($AUC > 0.95$), while the classifier based on the CpGs in the dominant regions had a relatively higher predictive ability than the classifier based on the CpGs in all of the regions.

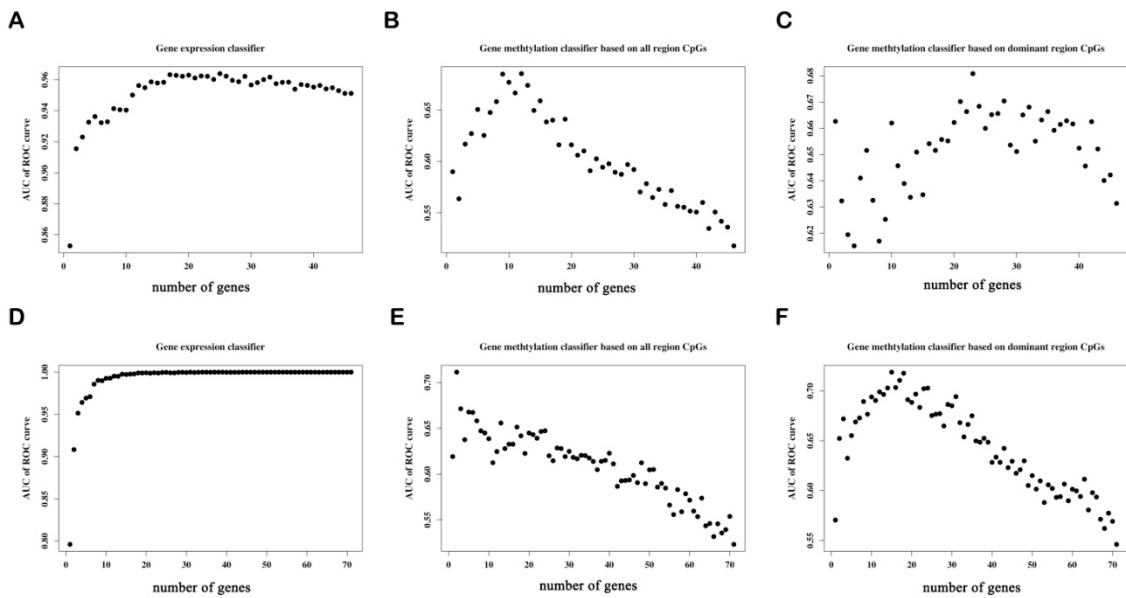


Figure 8. Scatter plots diagrammatizing the relationship between the prediction ability and the number of hypo-up genes and hyper-down genes in each classifier. Classifiers of the hypo-up genes: (A) The gene expression classifier, (B) the gene methylation classifier based on the CpGs in all of the regions, and (C) the gene methylation classifier based on the CpGs in the dominant regions. Classifiers of hyper-down genes: (D) The gene expression classifier, (E) the gene methylation classifier based on the CpGs in all of the regions, and (F) the gene methylation classifier based on the CpGs in the dominant regions. ROC, receiver operating characteristic; AUC, area under the curve; y-axis, AUC value of the ROC curve for the classifier; x-axis, the number of genes in the classifier.

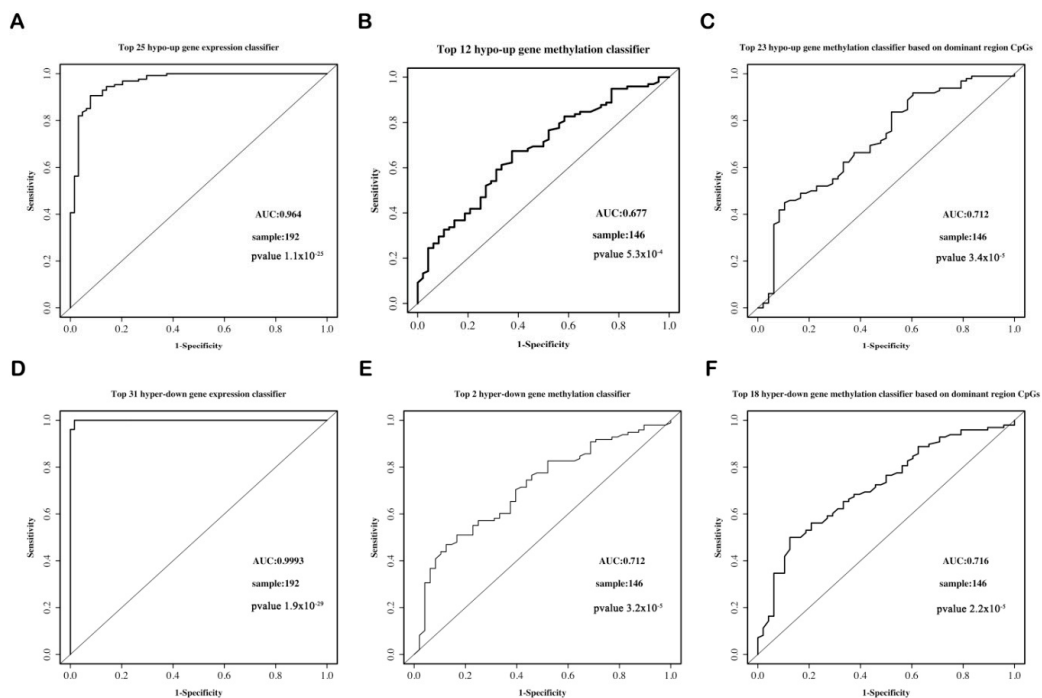


Figure 9. ROC curves for the hypo-up genes and hyper-down gene classifiers. (A) Top 25 hypo-up gene expression classifier. (B) Top 12 hypo-up gene methylation classifier based on the CpGs in all of the regions. (C) Top 23 hypo-up gene methylation classifier based on the CpGs in the dominant regions. (D) Top 31 hyper-down gene expression classifier. (E) Top 2 hyper-down gene methylation classifier based on the CpGs in all of the regions. (F) Top 18 hyper-down gene methylation classifier based on the CpGs in the dominant regions. ROC, receiver operating characteristic; AUC, area under the curve.

4. Discussion

MDD is a mental disorder with high morbidity, a complicated etiology, and a severe socioeconomic burden, lacking diagnostic biomarkers. By conducting bioinformatics analysis and mining of the GSE98793 gene expression dataset and the GSE113725 genome-wide methylation dataset, we obtained the following results. DEGs are mainly enriched in the inflammatory response-associated pathways, while DMGs are mainly enriched in the neurodevelopmental- and neuroplasticity-associated pathways. Through integrated analysis of the gene expression and DNA methylation data, 46 hypo-up genes and 71 hyper-down genes were identified. These genes are mainly involved in immune activation, synaptic development, and DNA repair. Classifiers based on the gene expression and DNA methylation in all regions, as well as in the different regions, were established by the random forest algorithm and the LOO cross-validation method. The results reveal that for both the hypo-up and hyper-down genes, the classifier based on the gene expression data exhibited the best predictive power, while the methylation classifier based on the CpGs in the dominant regions possessed a relatively higher predictive ability than the methylation classifier based on the CpGs in all regions.

Sigmund Freud wrote that “the complex of melancholia behaves like an open wound” [34]. Clinical and translational studies have shown that inflammatory responses are associated with the onset and maintenance of MDD. Glial cells, including microglia and astrocytes, are the primary immune mediators of the brain and respond accordingly to external stimuli. Proinflammatory (TNF- α and IL-1 β) and anti-inflammatory (IL-1, IL-10, and TGF β 1) factors are released under stress. Increasing levels of proinflammatory mediators such as IL-1, IL-2, IL-6, and TNF- α have been observed in patients with depression [35]. Herein, the GO enrichment analysis indicated that DEGs are associated with neutrophil-mediated immunity and the neutrophil activation involved in immune response, and the results are consistent with previous studies. A growing body of research shows that inflammation is closely related to depression [36], but the exact molecular mechanisms remain to be elucidated. Based on the results of our analysis, we hypothesize that inflammatory responses are involved in the progression of depression, and circulating inflammatory factors could be potential diagnostic biomarkers for MDD.

Interestingly, different from DEGs, DMGs are enriched in axonogenesis, neuron projection guidance, axon guidance, GO terms, and the MAPK signaling pathways. The postmortem and meta-analyses of magnetic resonance imaging studies indicate that hippocampal volume decreases in patients with depression [37,38]. There are two hypotheses for this phenomenon, namely, the neuroplasticity hypothesis and the neurogenesis hypothesis. The former suggests that stress induces the atrophy of mature neurons in the hippocampus, while the latter suggests that stress decreases the number of newborn neurons and neural precursor cells in the dentate gyrus of the hippocampus. As mentioned in the last paragraph, astrocytes and microglia participate in the inflammatory response in the brain, and they also secrete nutrients, such as BDNF, to nourish neurons. BDNF positively regulates nerve polymorphisms, and BDNF expression is decreased in the hippocampus of depressed patients [39,40]. Research suggests that the methylation level of the BDNF promoter region in MDD patients is increased, while the mRNA expression is decreased [41]. It has also been reported that IL-6 modulates synaptic plasticity [42]. We presume that there could be extremely complex crosstalk among inflammatory response, neuroplasticity, gene expression, and DNA methylation, but the molecular mechanisms remain a mystery.

TSS1500, TSS200, 5'UTR, and 1stExon are all related to transcriptional initiation and can be subsumed into the promoter region. Our results showed that DMGs have the highest distribution in the promoter region, and the gene body region also accounts for a large proportion. Basically, there is one CpG island (CGI) in every 10 base pairs in the human genome. However, the content of CGI surrounding the TSS of protein-coding genes is as high as 60%. The promoter region is the most studied region of DNA methylation, and it is considered that the CGI methylation of the promoter region is a hallmark of inhibiting gene expression.

Little research has been done on DNA methylation in the gene body region. Benefitting from the development of sequencing technology, we have a more complete understanding of genome-wide DNA methylation modification. Except for the promoter region, many CGIs are distributed in the gene body and intergenic region. Ehrlich et al. [43] found that brain tissue contains some of the highest levels of DNA methylation in the gene body. In the human brain, 16% of all CGIs are methylated, while 98% of the annotated 5' promoter regions are unmethylated and, surprisingly, CGI methylation in the gene body region is up to 34% [44]. Although the methylation function of the gene body is unclear, we speculate that methylation in the gene body region may be more dynamic than in the promoter region. Evidence suggests that the degree of gene body methylation in dividing cells is positively correlated with gene expression [45]. Unlike the methylation of the promoter region, which inhibits transcription initiation, gene body methylation does not prevent—and may even promote—transcription elongation [33]. Our findings provide new insight into research related to gene body methylation and depression.

We constructed three types of classifiers: A gene expression classifier, a methylation classifier based on the CpGs in all of the regions, and a methylation classifier based on the CpGs in the dominant regions for both the 46 hypo-up genes and the 71 hyper-down genes. The classifier based on the gene expression data exhibited satisfactory predictive ability, with an AUC > 0.95. The predictive ability of the methylation classifier based on the CpGs in the dominant regions was relatively better than that of the methylation classifier based on the CpGs in all of the regions. The relationship between DNA methylation and depression is a controversial topic. The results presented by genome-wide DNA methylation studies are multitudinous [20,21]. A detailed study of a certain region(s) or CpG(s) should better demonstrate the relationship between the two. A post-hoc investigation indicated that FKBP5 intron methylation has a negative correlation with transcription activation in MDD patients [46]. In a prospective analysis of major depressive disorder in adolescent girls, the authors found that all four significant CpGs in NR3C1 were in the gene body region: Two sites were located within a transcription factor-binding site (TFBS) region, one was in a region of open chromatin, and one site associated with an enhancer element [47]. Benjamard et al. [48] reported that in the promoter region of parvalbumin, methylation was significantly increased at CpG2 and decreased at CpG4 in the MDD group compared to the control group. Some alterations of CpGs are limited to specific gene phenotypes. In the SS genotype of 5-HTTLPR, depression is significantly involved with a decrease in methylation levels at CpG21, CpG25, and CpG26 [49]. Studies based on a certain region(s) or CpG(s) level should be the future research trend of the relationship between DNA methylation and MDD.

Although all of the classifiers demonstrated favorable predictive ability, the limitations cannot be ignored. First, the samples of gene expression and DNA methylation data came from different cohorts. Complicated diseases (such as MDD) involve molecular changes at multiple levels, such as at the genome, epigenome, and transcriptome levels. Researchers hope to systematically and comprehensively study the pathogenesis of diseases from multiple dimensions and perspectives. However, due to limited data sources and research funding, many studies have been conducted using datasets of similar disease models or similar research backgrounds. For example, Reference [26] integrated DNA methylation and transcriptome data and identified 85 hypo-up genes that could be potential diagnostic biomarkers for Parkinson's disease. The same integrated analysis was performed in Reference [50] to predict gastric cancer. Integrated analyses of other omics are also common in medical research, namely, multi-genome [51], and multi-transcriptome [52] analyses. Not only is this phenomenon observed in clinical studies, but also in botanical studies (e.g., multi-transcriptome in References [53,54], and multi-ChIP-seq in Reference [55]). Although our data did not match completely, we used the integrated analysis method of different omics to provide a new perspective and direction for depression. Second, to what extent changes in peripheral blood genes are associated with genes in the brain is unknown, and integrated analysis of peripheral blood samples and brain tissues is necessary. Finally,

further experimental validation will improve the credibility of genes in the classifiers as potential biomarkers for MDD.

5. Conclusions

For the first time, three types of classifiers (i.e., a gene expression classifier, a methylation classifier based on the CpGs in all of the regions, and a methylation classifier based on the CpGs in the dominant regions) were constructed and compared with one another on the basis of integrated analysis in MDD. The results showed that for the 46 hypo-up genes and the 71 hyper-down genes, the gene expression classifier presented the best predictive power, while the methylation classifier based on the CpGs in the dominant regions was relatively better than the methylation classifier based on the CpGs in all of the regions. Taken together, we identified a blood signature consisting of 46 hypo-up genes and 71 hyper-down genes, which may play a potential role in the diagnosis of MDD.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2073-4425/12/2/178/s1>: Table S1: Demographic and clinical features for GSE113725. Table S2: Demographic and clinical features for the GSK-HiTDiP and Janssen-BRC (GSE98793) case-control studies. Table S3: Gene symbols of 1056 differentially expressed genes (DEGs). Table S4: Gene symbols of 8313 differentially methylated genes (DMGs). Table S5: Pathways that the 46 hypo-up genes are involved in. Table S6: Pathways that the 71 hyper-down genes are involved in. Table S7: Relevant information about the methylation of CpGs in all regions of the 46 hypo-up genes. Table S8: Relevant information about the methylation of CpGs in the dominant regions of the 46 hypo-up genes. Table S9: Relevant information about the methylation of CpGs in all regions of the 71 hyper-down genes. Table S10: Relevant information about methylation of CpGs in the dominant regions of the 71 hyper-down genes.

Author Contributions: Y.X. and G.W. designed the study, conducted the transcriptional and methylation data analysis, and drafted the manuscript. L.X. and L.C. carried out integration of the methylation and gene expression data and constructed classifiers based on gene expression and DNA methylation using the random forest algorithm and the leave-one-out cross-validation method. Y.Z. and C.Z. were responsible for data screening and collection. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant numbers 81871072 and 82071523) and the Medical Science Advancement Program of Wuhan University (grant number TFLC2018001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Busch, Y.; Menke, A. Blood-based biomarkers predicting response to antidepressants. *J. Neural Transm.* **2019**, *126*, 47–63. [[CrossRef](#)] [[PubMed](#)]
2. Chirita, A.L.; Gheorman, V.; Bondari, D.; Rogoveanu, I. Current understanding of the neurobiology of major depressive disorder. *Rom. J. Morphol. Embryol.* **2015**, *56*, 651–658. [[PubMed](#)]
3. Kennis, M.; Gerritsen, L.; van Dalen, M.; Williams, A.; Cuijpers, P.; Bockting, C. Prospective biomarkers of major depressive disorder: A systematic review and meta-analysis. *Mol. Psychiatry* **2019**, *25*, 321–338. [[CrossRef](#)]
4. Hasin, D.S.; Sarvet, A.L.; Meyers, J.L.; Saha, T.D.; Ruan, W.J.; Stohl, M.; Grant, B.F. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry* **2018**, *75*, 336–346. [[CrossRef](#)] [[PubMed](#)]
5. Nemeroff, C.B. The burden of severe depression: A review of diagnostic challenges and treatment alternatives. *J. Psychiatr. Res.* **2007**, *41*, 189–206. [[CrossRef](#)]
6. Siegle, G.J.; Carter, C.S.; Thase, M.E. Use of fMRI to predict recovery from unipolar depression with cognitive behavior therapy. *Am. J. Psychiatry* **2006**, *163*, 735–738. [[CrossRef](#)]
7. Hepgul, N.; Cattaneo, A.; Zunszain, P.A.; Pariante, C.M. Depression pathogenesis and treatment: What can we learn from blood mRNA expression? *BMC Med.* **2013**, *11*, 28. [[CrossRef](#)]

8. Tavakolizadeh, J.; Roshanaei, K.; Salmaninejad, A.; Yari, R.; Nahand, J.S.; Sarkarizi, H.K.; Mousavi, S.M.; Salarinia, R.; Rahmati, M.; Mousavi, S.F.; et al. MicroRNAs and exosomes in depression: Potential diagnostic biomarkers. *J. Cell Biochem.* **2018**, *119*, 3783–3797. [[CrossRef](#)]
9. Chamberlain, S.R.; Cavanagh, J.; de Boer, P.; Mondelli, V.; Jones, D.N.C.; Drevets, W.C.; Cowen, P.J.; Harrison, N.A.; Pointon, L.; Pariante, C.M.; et al. Treatment-resistant depression and peripheral C-reactive protein. *Br. J. Psychiatry* **2019**, *214*, 11–19. [[CrossRef](#)]
10. Lohoff, F.W. Overview of the genetics of major depressive disorder. *Curr. Psychiatry Rep.* **2010**, *12*, 539–546. [[CrossRef](#)]
11. Sullivan, P.F.; Neale, M.C.; Kendler, K.S. Genetic epidemiology of major depression: Review and meta-analysis. *Am. J. Psychiatry* **2000**, *157*, 1552–1562. [[CrossRef](#)] [[PubMed](#)]
12. Sun, H.; Kennedy, P.J.; Nestler, E.J. Epigenetics of the depressed brain: Role of histone acetylation and methylation. *Neuropsychopharmacology* **2013**, *38*, 124–137. [[CrossRef](#)] [[PubMed](#)]
13. Jaenisch, R.; Bird, A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat. Genet.* **2003**, *33*, 245–254. [[CrossRef](#)] [[PubMed](#)]
14. Li, M.; D’Arcy, C.; Li, X.; Zhang, T.; Joobar, R.; Meng, X. What do DNA methylation studies tell us about depression? A systematic review. *Transl. Psychiatry* **2019**, *9*, 68–81. [[CrossRef](#)] [[PubMed](#)]
15. Klengel, T.; Pape, J.; Binder, E.B.; Mehta, D. The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology* **2014**, *80*, 115–132. [[CrossRef](#)]
16. Menke, A.; Binder, E.B. Epigenetic alterations in depression and antidepressant treatment. *Dialogues Clin. Neurosci.* **2014**, *16*, 395–404.
17. Lisoway, A.J.; Zai, C.C.; Tiwari, A.K.; Kennedy, J.L. DNA methylation and clinical response to antidepressant medication in major depressive disorder: A review and recommendations. *Neurosci. Lett.* **2017**, *669*, 14–23. [[CrossRef](#)]
18. Hüls, A.; Robins, C.; Conneely, K.N.; De Jager, P.L.; Bennett, D.A.; Epstein, M.P.; Wingo, T.S.; Wingo, A.P. Association between DNA methylation levels in brain tissue and late-life depression in community-based participants. *Transl Psychiatry.* **2020**, *10*, 262–271. [[CrossRef](#)]
19. Clark, S.L.; Hattab, M.W.; Chan, R.F.; Shabalin, A.A.; Han, L.K.M.; Zhao, M.; Smit, J.H.; Jansen, R.; Milaneschi, Y.; Xie, L.Y.; et al. A methylation study of long-term depression risk. *Mol. Psychiatry* **2020**, *25*, 1334–1343. [[CrossRef](#)]
20. Tseng, P.T.; Lin, P.Y.; Lee, Y.; Hung, C.F.; Lung, F.W.; Chen, C.S.; Chong, M. Age-associated decrease in global DNA methylation in patients with major depression. *Neuropsychiatr. Dis. Treat.* **2014**, *10*, 2105–2114.
21. Uddin, M.; Koenen, K.C.; Aiello, A.E.; Wildman, D.E.; de los Santos, R.; Galea, S. Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample. *Psychol. Med.* **2011**, *41*, 997–1007. [[CrossRef](#)] [[PubMed](#)]
22. Crawford, B.; Craig, Z.; Mansell, G.; White, I.; Smith, A.; Spaul, S.; Imm, J.; Hannon, E.; Wood, A.; Yaghootkar, H.; et al. DNA methylation and inflammation marker profiles associated with a history of depression. *Hum. Mol. Genet.* **2018**, *27*, 2840–2850. [[CrossRef](#)] [[PubMed](#)]
23. Leday, G.G.R.; Vertes, 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](#)]
24. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)]
25. Weinhold, L.; Wahl, S.; Pechlivanis, S.; Hoffmann, P.; Schmid, M. A statistical model for the analysis of Beta values in DNA methylation studies. *BMC Bioinform.* **2016**, *17*, 480–490. [[CrossRef](#)]
26. Wang, C.; Chen, L.; Yang, Y.; Zhang, M.; Wong, G. Identification of potential blood biomarkers for Parkinson’s disease by gene expression and DNA methylation data integration analysis. *Clin. Epigenet.* **2019**, *11*, 24–39. [[CrossRef](#)]
27. Xu, T.; Li, B.; Zhao, M.; Szulwach, K.E.; Street, R.C.; Lin, L.; Yao, B.; Zhang, F.; Jin, P.; Wu, H.; et al. Base-resolution methylation patterns accurately predict transcription factor bindings in vivo. *Nucleic Acids Res.* **2015**, *43*, 2757–2766. [[CrossRef](#)]
28. Rajagopal, N.; Xie, W.; Li, Y.; Wagner, U.; Wang, W.; Stamatoyannopoulos, J.; Ernst, J.; Kellis, M.; Ren, B. RFECFS: A random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput. Biol.* **2013**, *9*, e1002968. [[CrossRef](#)]
29. Bufalino, C.; Hepgul, N.; Aguglia, E.; Pariante, C.M. The role of immune genes in the association between depression and inflammation: A review of recent clinical studies. *Brain Behav. Immun.* **2013**, *31*, 31–47. [[CrossRef](#)]
30. Debnath, M.; Doyle, K.M.; Langan, C.; McDonald, C.; Leonard, B.; Cannon, D.M. Recent advances in psychoneuroimmunology: Inflammation in psychiatric disorders. *Transl. Neurosci.* **2011**, *2*, 121–137. [[CrossRef](#)]
31. Messay, B.; Lim, A.; Marsland, A.L. Current understanding of the bi-directional relationship of major depression with inflammation. *Biol. Mood Anxiety Disor.* **2012**, *2*, 4–7. [[CrossRef](#)] [[PubMed](#)]
32. Hiles, S.A.; Bake, R.A.L.; de Malmanche, T.; Attia, J. A meta-analysis of differences in IL-6 and IL-10 between people with and without depression: Exploring the causes of heterogeneity. *Brain Behav. Immun.* **2012**, *2626*, 1180–1188. [[CrossRef](#)] [[PubMed](#)]
33. Jones, P.A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **2012**, *13*, 484–492. [[CrossRef](#)] [[PubMed](#)]
34. Freud, S. Mourning and Melancholia. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud*; Freud, S., Strachey, J., Eds.; Hogarth Press: London, UK, 1957; Volume 14, pp. 243–258.
35. Dowlati, Y.; Herrmann, N.; Swardfager, W.; Liu, H.; Sham, L.; Reim, E.K.; Lanctôt, K.L. A meta-analysis of cytokines in major depression. *Biol. Psychiatry* **2010**, *67*, 446–457. [[CrossRef](#)]

36. Slavich, G.M.; Irwin, M.R. From stress to inflammation and major depressive disorder: A social signal transduction theory of depression. *Psychol. Bull.* **2014**, *140*, 774–815. [[CrossRef](#)]
37. Kempton, M.J.; Salvador, Z.; Munafo, M.R.; Geddes, J.R.; Simmons, A.; Frangou, S.; Williams, S.C. Structural neuroimaging studies in major depressive disorder. Meta-analysis and comparison with bipolar disorder. *Arch. Gen. Psychiatry* **2011**, *68*, 675–690. [[CrossRef](#)]
38. Cobb, J.A.; Simpson, J.; Mahajan, G.J.; Overholser, J.C.; Jurjus, G.; Dieter, L.; Herbst, N.; May, W.; Rajkowska, G.; Stockmeier, C.A. Hippocampal volume and total cell numbers in major depressive disorder. *J. Psychiatr. Res.* **2013**, *47*, 299–306. [[CrossRef](#)]
39. Alfonso, J.; Frick, L.R.; Silberman, D.M.; Palumbo, M.L.; Gerano, A.M.; Frasch, A.C. Regulation of hippocampal gene expression is conserved in two species subjected to different stressors and antidepressant treatment. *Biol. Psychiatry* **2006**, *59*, 244–251. [[CrossRef](#)]
40. Patel, M.N.; McNamara, J.O. Selective enhancement of axonal branching of cultured dentate gyrus neurons by neurotrophic factors. *Neuroscience* **1995**, *69*, 763–770. [[CrossRef](#)]
41. Januar, V.; Ancelin, M.L.; Ritchie, K.; Saffery, R.; Ryan, J. BDNF promoter methylation and genetic variation in late-life depression. *Transl. Psychiatry* **2015**, *5*, e619. [[CrossRef](#)]
42. Wang, J.; Hodes, G.E.; Zhang, H.; Zhang, S.; Zhao, W.; Golden, S.A.; Bi, W.; Menard, C.; Kana, V.; Leboeuf, M.; et al. Epigenetic modulation of inflammation and synaptic plasticity promotes resilience against stress in mice. *Nat. Commun.* **2018**, *9*, 477–490. [[CrossRef](#)] [[PubMed](#)]
43. Ehrlich, M.; Sosa, M.A.G.; Huang, L.H.; Midgett, R.M.; Kuo, K.C.; McCune, R.A.; Gehrke, E.C. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.* **1982**, *10*, 2709–2721. [[CrossRef](#)]
44. Maunakea, A.K.; Nagarajan, R.P.; Bilenky, M.; Ballinger, T.J.; D’Souza, C.; Fouse, S.D.; Johnson, B.E.; Hong, C.; Nielsen, C.; Zhao, Y.; et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **2010**, *466*, 253–257. [[CrossRef](#)] [[PubMed](#)]
45. Aran, D.; Toperoff, G.; Rosenberg, M.; Hellman, A. Replication timing-related and gene body-specific methylation of active human genes. *Hum. Mol. Genet.* **2011**, *20*, 670–680. [[CrossRef](#)] [[PubMed](#)]
46. Tozzi, L.; Farrell, C.; Booij, L.; Doolin, K.; Nemoda, Z.; Szyf, M.; Pomares, F.B.; Chiarella, J.; O’Keane, V.; Frodl, T. Epigenetic changes of FKBP5 as a link connecting genetic and environmental risk factors with structural and functional brain changes in major depression. *Neuropsychopharmacology* **2018**, *43*, 1138–1145. [[CrossRef](#)] [[PubMed](#)]
47. Humphreys, K.L.; Moore, S.R.; Davis, E.G.; MacIsaac, J.L.; Lin, D.T.S.; Kobor, M.S.; Gotlib, I.H. DNA methylation of HPA-axis genes and the onset of major depressive disorder in adolescent girls: A prospective analysis. *Transl. Psychiatry* **2019**, *9*, 245–254. [[CrossRef](#)] [[PubMed](#)]
48. Sukjai, B.T.; Suttajit, S.; Thanoi, S.; Dalton, C.F.; Reynolds, G.P.; Nudmamud-Thanoi, S. Parvalbumin promoter methylation altered in major depressive disorder. *Int. J. Med. Sci.* **2019**, *16*, 1207–1214. [[CrossRef](#)]
49. Lam, D.; Ancelin, M.L.; Ritchie, K.; Poli, R.F.; Saffery, R.; Ryan, J. Genotype-dependent associations between serotonin transporter gene (SLC6A4) DNA methylation and late-life depression. *BMC Psychiatry* **2018**, *18*, 282. [[CrossRef](#)]
50. Peng, Y.; Wu, Q.; Wang, L.; Wang, H.; Yin, F. A DNA methylation signature to improve survival prediction of gastric cancer. *Clin. Epigenetics.* **2020**, *12*, 15–30. [[CrossRef](#)]
51. Griffith, O.L.; Pepin, F.; Enache, O.M.; Heiser, L.M.; Collisson, E.A.; Spellman, P.T.; Gray, J.W. A robust prognostic signature for hormone-positive node-negative breast cancer. *Genome Med.* **2013**, *5*, 92–105. [[CrossRef](#)]
52. Zeng, D.; Li, M.; Zhou, R.; Zhang, J.; Sun, H.; Shi, M.; Bin, J.; Liao, Y.; Rao, J.; Liao, W. Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol. Res.* **2019**, *7*, 737–750. [[CrossRef](#)] [[PubMed](#)]
53. Song, Q.; Lee, J.; Akter, S.; Rogers, M.; Grene, R.; Li, S. Prediction of condition-specific regulatory genes using machine learning. *Nucleic Acids Res.* **2020**, *48*, e62. [[CrossRef](#)] [[PubMed](#)]
54. Shaik, R.; Ramakrishna, W. Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. *Plant. Physiol.* **2014**, *164*, 481–495. [[CrossRef](#)] [[PubMed](#)]
55. Chow, C.N.; Lee, T.Y.; Hung, Y.C.; Li, G.Z.; Tseng, K.C.; Liu, Y.H.; Kuo, P.L.; Zheng, H.Q.; Chang, W.C. PlantPAN3.0: A new and updated resource for reconstructing transcriptional regulatory networks from ChIP-seq experiments in plants. *Nucleic Acids Res.* **2019**, *47*, D1155–D1163. [[CrossRef](#)]