

Network modules linking expression and methylation in prefrontal cortex of schizophrenia

Dongdong Lin¹, Jiayu Chen¹, Nora Perrone-Bizzozero^{2,4}, Jing Sui¹, Vince D. Calhoun^{1,2,3,*}, Jingyu Liu^{1,3,*}

¹The Mind Research Network and Lovelace Biomedical and Environmental Research Institute, Albuquerque, NM, 87106, USA

²Dept. of Neurosciences, University of New Mexico, Albuquerque, NM, 87131, USA

³Dept. of Electronic and Computer Engineering, University of New Mexico, Albuquerque, NM, 87131, USA

⁴Dept. of Psychiatry, University of New Mexico, Albuquerque, NM, 87131, USA

*these authors contributed equally to the work

Abstract

Tremendous work has demonstrated the critical roles of genetics, epigenetics as well as their interplay in brain transcriptional regulations in the pathology of schizophrenia (SCZ). There is great success currently in the dissection of the genetic components underlying risk-conferring transcriptomic networks. However, the study of regulating effect of epigenetics, as a modulator of environmental factors, in the etiopathogenesis of SCZ still faces many challenges. In this work we investigated DNA methylation and gene expression from the dorsolateral prefrontal cortex (DLPFC) region of schizophrenia patients and healthy controls using weighted correlation network approaches. We identified and replicated two expression and two methylation modules significantly associated with SCZ. Among them, one pair of expression and methylation modules were significantly overlapped in the module genes which were enriched in astrocyte-associated functional pathways, and specifically expressed in astrocytes. Another two linked expression-methylation module pairs were involved aging process with module genes mostly related to oligodendrocyte development and myelination, and specially expressed in oligodendrocytes. Further examination of underlying quantitative trait loci (QTLs) showed significant enrichment in genetic risk of most psychiatric disorders for expression QTLs but not for methylation QTLs. These results support the coherence between methylation and gene expression in a network level, and suggest a combinatorial effect of genetics and epigenetics in regulating gene expression networks specific to glia cells in relation with SCZ and aging process.

Keywords: DNA methylation, gene expression, brain postmortem, and schizophrenia

Introduction

Schizophrenia (SCZ) is known to be a highly heritable and developmental neuropsychiatric disorder, which has around 1% prevalence worldwide [1, 2]. In previous neuroimaging studies, SCZ has been characterized by brain-wide gray matter reduction [3], functional disconnectivity [4, 5] and myelination decrease [6, 7] along with significant impairments in cognition [8], and presence of positive and negative symptoms. These brain abnormalities are thought to be mediated by complex molecular processes involving neuronal cell development, differentiation and death, even during the early stage of brain development [9]. Despite substantial efforts on exploring the underlying molecular mechanisms through advanced genomics, epigenetics and transcriptomics techniques, there are still many challenges given the complex interplay among these factors for the etiology of SCZ.

Brain transcriptomics studies have demonstrated substantial transcriptional alterations in the frontal cortex, cerebellum or hippocampus of SCZ [10]. Keeping gene expression in balance is critical in maintaining neuronal function, as well as their synaptic interactions. Genetics is a major factor in regulating gene expression and functional pathways [11, 12]. Recent landmark work based on dorsolateral prefrontal cortex (DLPFC) postmortem tissue RNA sequencing has reported 20~50% of SCZ risk loci [13] showing strong cis-effect to their nearby genes' expression, further elaborating the liability of these risk variants in etiology of SCZ [14, 15]. Co-expression network analyses of brain expression data have identified some modules highly associated with SCZ harboring the SCZ risk loci with strong cis-acting effects [15, 16]. However, some co-expression modules show discrepancy in relation with SCZ and the polygenic risk score for SCZ. Several SCZ-related modules fail to be significantly enriched in SCZ genetic susceptibility [14]. In addition, differentially expressed genes could have diverse patterns among

brain regions [17], neural cells [18, 19] and neural developmental stages [20]. Collectively, these findings suggest the influence of complex interplay of genetics and environmental factors in determining the process of gene expression.

Epigenetics can mediate gene by environment effects in modifying how genes are structured and expressed. DNA methylation is an epigenetic modification widely studied in psychiatric disorders such as SCZ [21, 22]. It changes the genome's response to transcriptional factors by attaching a methyl group in DNA sequence (mostly on cytosine site as CpG). The influence of DNA methylation in gene transcription can be stable and heritable across cell generations, and also reversible according to external condition changes [23]. The dynamics of DNA methylation plays a vital role in the pathogenesis of SCZ, especially in neuronal diversity, plasticity and neurogenesis [24, 25]. Besides environmental and developmental effects, DNA methylation can also be influenced by sequence variants (e.g., genotype variation or specific allele on a locus), which represents a methylation QTL (meQTL). Previous studies have demonstrated prevalent meQTL effects across genome, especially during the early neural development, and their relation to multiple psychiatric disorders including SCZ [26-28]. The study for common and tissue specific meQTL effects across brain and peripheral tissues shows that cross-tissue meQTLs are more likely to reside at regulatory elements (e.g., enhancer) of transcriptional genes and also highly enriched in eQTLs [29], suggesting the role of DNA methylation in mediating genetic effects on gene expression.

Although DNA methylation plays a very important role in the control of gene expression by modulating both genetics and environmental factors, there are currently limited number of studies directly investigating the relationship between DNA methylation and gene expression in the human brain. By leveraging public datasets of methylation and gene expression from brain

DLPFC tissues, we performed a comprehensive network analysis with an aim to characterize the relationship between them. Given the co-regulation among genes in expression and the clustering of methylation in functional pathways, we sought to identify the relationship between functional transcriptional modules and epigenetic modules in the context of the human interactome. We applied network analyses to construct expression and methylation networks and prioritize them into multiple sub-network modules, followed by a systematic characterization of both types of modules in terms of their associations with SCZ and aging, neuronal cell specification, CpG-expression co-localization and their underlying genetic effects. Using these analyses, we were able to identify links between gene expression and methylation at the module level.

Materials and Methods

Gene expression data

The data were downloaded from dbGaP (Accession: phs000979.v1.p1). Briefly, brain postmortem tissues of DLPFC gray matter region (Brodmann area 46) from 546 subjects including SCZ and HCs were used for RNA extraction and gene expression assays. Further information about tissue dissection, clinical characterization, neuropathological screening, and toxicological analyses was described in [30]. SCZ patients were assessed for schizophrenia or schizoaffective disorder based on DSM-IV lifetime Axis I. HCs were included if they had no history of psychological or psychiatric problems and negative toxicology results.

Gene expression was assayed by an Illumina whole-genome HT-12-V4 expression chip covering 47323 probes. The transcript level of each probe was passed through quality control steps by the 'limma' R package [31], including correction for background level using negative control probes, quantile normalization using both negative and positive controls, and \log_2

transformation on the transcription intensities. Probes were removed if they were not significantly expressed (detection p-value > 0.01) in at least 90% subjects as applied in [32]. Subjects were excluded if they had age information missing or were less than 16 years old, or had poor RNA quality (RNA integrity number: $RIN < 6.5$) or less than 10% of probes significantly expressed (quality detection p-value < 0.01). After these quality control steps, we had 20,370 probes from 419 subjects (253 HCs and 250 SCZ) left for analyses. Batch effects were corrected for each probe using a parametric Bayes framework implemented in the ‘combat’ function [33] in an R package ‘SVA’[34]. Confounding factors including pH, PMI (post-mortem interval) and RIN were further regressed out by a linear model prior to main analyses.

DNA methylation data

The data were from NCBI GEO database (GEO database: GSE74193). Briefly DNAs from DLPFC region (BA9/46) of 244 subjects (108 SCZ and 136 HCs) with age greater than 16 years old were assayed using Illumina Infinium Methylation450k assay, covering 485,512 CpG sites. More details about the criteria for patient and controls recruitment and diagnosis can be seen in [26].

DNA methylation data went through a series of quality control steps performed by R package ‘minfi’ [35] as applied in [26, 36]. After applying quantile-based normalization to both methylated and unmethylated signals on each site, we calculated the beta values for subsequent analysis. CpGs were removed if they 1) coincided with SNPs or at single base extension [37], 2) located in non-specific probes [38], 3) contained more than 1% missing values (methylation values with detection $p > 0.05$ were treated as missing values), or 4) were located on sex chromosomes. The remaining missing beta values were further imputed by KNN method as used

in [39]. 377,698 CpGs were kept after preprocessing. Further CpG removal was applied if CpGs standard deviation (sd) was less than the measurement error standard deviation ($sd = 0.047$) estimated by 78 test-retest samples, resulting 61853 CpGs for analyses. Batch effects were then corrected for each CpG using the same steps as with expression data. Neuronal cell type proportions and the top four principle components of negative control methylation as estimated by [26] were regressed out for subsequent analysis.

Network analysis

After preprocessing of both expression and methylation data, we constructed co-expression and co-methylation modules separately by using software WGCNA [40]. In brief, for each dataset the adjacency matrix was calculated by a power of 6 of the correlation matrix among nodes (i.e., expression probes and CpG) to have scale-free topology larger than 0.85, from which topology overlap matrix (TOM) was derived to measure connection similarity among nodes—the overlap between any two nodes in terms of the extent they were connected to the same other nodes in the network. Through the TOM matrix, an unsigned co-methylation/expression network was constructed and densely interconnected CpGs or expression probes were clustered into modules. Module eigengenes (ME), the first principal component of methylation or expression matrix in a module, were computed and tested for the association with SZ diagnosis, controlling for age, race and sex. Within a module, the correlation of each expression probe or CpG with ME was computed as the measure of their module membership (MM) [41], indicating how close a CpG or probe relates to the module. Each CpG or probe's association with SZ diagnosis was also computed as group significance (GS), from which the correlation between MM and GS on each

CpG or probe in the module was tested. The top CpGs or probes with both high MM and GS values were utilized to represent the module for demonstration.

Overlap test between expression and methylation modules

After prioritizing probes or CpGs into different expression or methylation modules respectively, we assigned them with ensemble IDs based on genome assembly GRCh38 using R package 'BioMart' [42]. We only annotated the methylation sites to a gene and ensemble ID if it was located within the gene or nearby the gene transcription start site (TSS) (distance < 5k bps). For each pair of expression and methylation modules, we tested the significance of ensemble ID overlap by both two-sided Fisher's 2×2 table exact test and permutation test. In Fisher's exact test, given an expression module, we compared the odd of shared ensemble IDs with a methylation module to the odd of shared ensemble IDs with the other methylation modules. To account for the cases that some methylation sites were not annotated to any genes and the number of methylation sites varied by modules, we also applied a permutation method to reduce these influences in the test. For a specific expression module, the proportion of overlapped ensemble ID from a methylation module was taken as an observation. Then we generated null distribution by randomly sampling 10^5 sets of methylation sites with the same module size from the whole methylation set. For each randomly selected methylation set, the proportion of ensemble ID included in the expression module was calculated. The percentage of sampled methylation sets having higher proportion than the observed proportion was taken as empirical p-value, denoted by P_perm. The significance levels from both methods were used in determining the significance of overlap between expression and methylation modules.

Replication analyses

We used independent expression and methylation data sets to validate the findings from discovery data from three aspects: module's association with phenotypes, module's preservation, and probe-level associations of expression and methylation from linked modules.

Expression replication dataset 1(ExpRep1; GSE36192): postmortem samples from frontal cortex of 455 neurologically normal Caucasian subjects were assayed by HumanHT-12_v3 Expression BeadChips (48803 probes). More details about expression profiling were described in [43, 44]. After applying the same quality control and preprocessing (batch correction) as in discovery data, there were 343 subjects (age > 16) with 18,196 probes left for replication.

Expression replication dataset2(ExpRep2; GSE21138): postmortem tissues from DLPFC (BA 46) of 30 SCZ patients and 29 age- and sex-matched HCs [45] assayed by Affymetrix Human Genome U133 Plus 2.0 Array, covering 54675 probes, followed by normalization using dChip statistical model (DNA-Chip Analyzer), and log₂ transformation as applied in [45] and batch correction by 'SVA'[34]. Probes with present calls in at least one sample were left, resulting in 30,061 probes from 58 (30 SCZ and 28 HCs) subjects for analysis. Probes were assigned by gene symbols and matched with those from discovery expression data at gene level for replication.

Methylation replication dataset1(MethyRep1; GSE61380): 33 post-mortem brain DLPFC (BA9) samples (18 SCZ and 15 HCs) were obtained from Douglas Bell-Canada brain bank with more information about participant recruitment and diagnosis criteria introduced in [46]. DNA methylation was profiled using the Illumina Infinium HumanMethylation450K BeadChip and went through the same normalization, quality control, and imputation process as applied in discovery data to have 368,998 methylation sites remained for validation. Batch correction and

five neural cell type proportions estimation were also applied using the same algorithms as used in discovery data analysis.

Methylation replication dataset2 (MethyRep2; GSE36194): corresponding to expression replication dataset 1, methylation data from the same tissue of same subjects were profiled using with the Illumina Infinium HumanMethylation27K beadchip. The methylation data were preprocessed and quality controlled through the same steps as in discovery data analysis, resulting 26424 CpGs from 330 subjects for validation analysis.

Replication for modular's disease and age relatedness: For those modules with ME vectors significantly associated with diagnosis, we validated the associations in the independent datasets. For expression modules, we used *ExpRep2* to select the probes from the same genes as the probes in the module and calculated the first PC of expression data for testing their associations with disease and age. A similar validation was applied for each significant methylation modules in *MethyRep1*. Replication significance level was set as 0.05.

Replication for module preservation: Three different measures were applied to validate if the properties of selected modules were preserved across different datasets. First, a local clustering coefficient is a normalized average number of all triangle connections associated with each node. The generalized clustering coefficient (GCC) by [47] for weighted network, accounting for weight information in each edge, was used to reflect the presence of the module structure and mean prevalence of connectivity around the nodes. A permutation test was applied to test the significance of GCC in replicated datasets. In each of 10^5 runs, we randomly selected features (expression probes or methylation sites) from the entire set and calculated GCC to build the null distribution. P-value can be derived by the proportion of runs with larger GCC than the observed

GCC. Second, a composite preservation statistic was previously proposed to combine the measures of module density preservation (i.e. to test if the connections in the module remain in an independent dataset) and module connectivity preservation (to test if the connectivity pattern is similar to that in the independent dataset) and tested in [48]. A permutation was performed to estimate the mean and variance of each measure (i.e., module density and module connectivity) under the null hypothesis that there was no preservation of module in the measures, followed by the standardization of each observed measure to the mean and variance under null distribution, denoted by Z_{density} and $Z_{\text{connectivity}}$, respectively. The composite preservation statistic, denoted by Z_{summary} , was defined as the average of Z_{density} and $Z_{\text{connectivity}}$. $Z_{\text{summary}} < 2$ indicates no preservation, $2 < Z_{\text{summary}} < 10$ indicates weak to moderate evidence of preservation, while $Z_{\text{summary}} > 10$ indicates strong evidence [48]. The p-value of the composite statistic was the combination of p values of module density and module connectivity.

Probe-level expression-methylation relation

To test directly regulation relation between expression and methylation at probe level for the linked modules, we leveraged the data from *ExpRep1* and *MethyRep2* that were collected from the same samples, and tested pair-wise associations of expression probes and methylation sites if they are located within 500k bps, using linear regression controlling for covariates age, sex and PMI.

Neural cell type and functional enrichment

To understand the cell specificity of each expression and methylation modules, we tested their enrichment in the genes expressed specifically in five cell types (neuron, astrocytes, microglia,

endothelia cells, and oligodendrocytes using the pSI R package [49] with Fisher's exact test threshold as 0.05 and false discovery rate (FDR) as 0.05. The purified cell expression data from GSE73721 were used for the test [50]. The overlap of our modules with those cell-specific modules identified in a recent transcriptional study for psychiatric disorders [51] was also evaluated by Fisher's exact test. In addition, methylation data of both neuron and glia cells from postmortem prefrontal tissues of 29 control subjects (GSE41826 [52]) were used to extract two sets of CpGs showing significant hyper-methylation in neuron (denoted by "neuron-up") and glia cells (denoted by "glia-up"), respectively ($p < 1 \times 10^{-10}$). These two sets of CpGs were then used to test the CpG overlap with identified methylation modules. To understand the biological function of each module, we performed a gene ontology (GO) and KEGG enrichment analysis on the genes in the module using the web tool Webgestalt (<http://www.webgestalt.org/option.php>).

Psychiatric disorders GWAS risk loci enrichment

Public meQTL database (GSE74193) and eQTL databases (GTEx: <https://gtexportal.org/home/>; BrainSeq: <http://eqtl.brainseq.org/phase1/devel>) from brain DLPFC tissue were applied to access the genetic effects underlying expression and methylation networks. To test if the QTLs from target modules were significantly enriched in genetic risk loci of psychiatric disorders compared to the QTLs from the other modules, we adopted both Fisher's exact test and permutation test as applied in [29]. Specifically, we firstly ran linkage disequilibrium (LD) pruning supervised by GWAS risk loci on the whole QTL set with $r^2 > 0.7$ using LD structure from 1000 genome project EUR group as reference. After pruning, the proportion of QTLs from the target module showing GWAS risk was taken as the observation. Then we generated null distribution by randomly sampling 10^5 sets of QTLs from the whole pruned QTL set with the same number of QTLs and

similar minor allele frequency distribution as those from the target module. Empirical p-value was computed by the percentage of sampled QTL sets having higher proportion than the observed proportion, denoted by P_{perm} . Fisher's exact test was used to test the odds ratio of risk loci in target module QTLs and the QTLs from the other modules, denoted by P_{Fisher} .

Results

Co-expression modules and their associations with SZ

We clustered expression probes into 22 modules using WGCNA software with the dendrogram plot as shown in Fig.1A. The eigengenes of each module were tested for association with multiple traits as shown in Fig.1B. Two modules colored by magenta (including 276 probes) and yellow (including 1098 probes) were identified to have significant associations with SCZ diagnosis (yellow: $t\text{-stat} = 4.1$, $p = 2 \times 10^{-3}$; magenta: $t\text{-stat} = 3.55$, $p = 9.7 \times 10^{-3}$) after controlling for covariates and correction for multiple tests by false positive rate (FDR). For each SCZ related module, we tested the correlation between module membership and SCZ group significance for all probes within the module, as shown in Fig.S1 A-B (Supplementary file 1). It can be seen that the correlation is significant in both modules with values of 0.53 and 0.28, respectively, showing that the probes with higher disease association tend to be more important in representing the module.

Independent validations were applied to verify both SCZ-related modules in two datasets (*ExpRep1* and *ExpRep2*). In *ExpRep1* dataset, 90.8% of expression probes from the yellow module and 88.8% probes from the magenta module were identified for replication. By permutation tests, we found that both modules were well preserved with significantly larger GCC

than random selection of probes ($p < 1 \times 10^{-4}$, Fig.S3 A-B) and higher Z_{summary} statistics (yellow module: $Z = 38$; magenta module: $Z = 18$, Fig.S1 C). As recommended in [48], if $Z_{\text{summary}} > 10$, there is strong evidence that the module is preserved. In *ExpRep2* dataset, 72.6% genes from the yellow module and 73.3% genes from the magenta module were identified. The expressions of those probes from the matched genes were used to calculate the first eigenvector for each module as module eigengene, followed by a test for schizophrenia association. We replicated significant disease associations of both modules with the same direction as identified in the discovery dataset (yellow module: $t\text{-stat} = 2.54$, $p = 0.014$; magenta module: $t\text{-stat} = 2.68$, $p = 9.6 \times 10^{-3}$). In addition, both modules consistently showed significantly larger GCC in *ExpRep2* data ($p < 1 \times 10^{-4}$, Fig.S3 C-D).

Several co-expression modules were significantly enriched in different neural cell types. For example, yellow expression module contains the genes mostly expressed in astrocytes while blue module is more specifically expressed in oligodendrocytes as shown in Fig.1C. Gene ontology enrichment analysis identified significant enrichment of genes from the yellow module in the pathways mainly involved the cell growth, migration, differentiation, movement, and organ development and morphogenesis as shown in Fig.1D ($FDR < 0.05$). Interestingly, they were also highly enriched in some neurodevelopment pathways such as gliogenesis, glial and astrocyte cell differentiation and central nervous system development. From Fig.1E, it can be seen that genes in the magenta module were mostly involved in the inflammatory response to external stimulus (e.g., cytokine and lipopolysaccharide), and were more specifically expressed in endothelial and microglia cells. Besides these modules, several others were also found to have significant enrichment in the cell-specific co-expression modules reported in a previous study [51], as shown in Fig.S4.

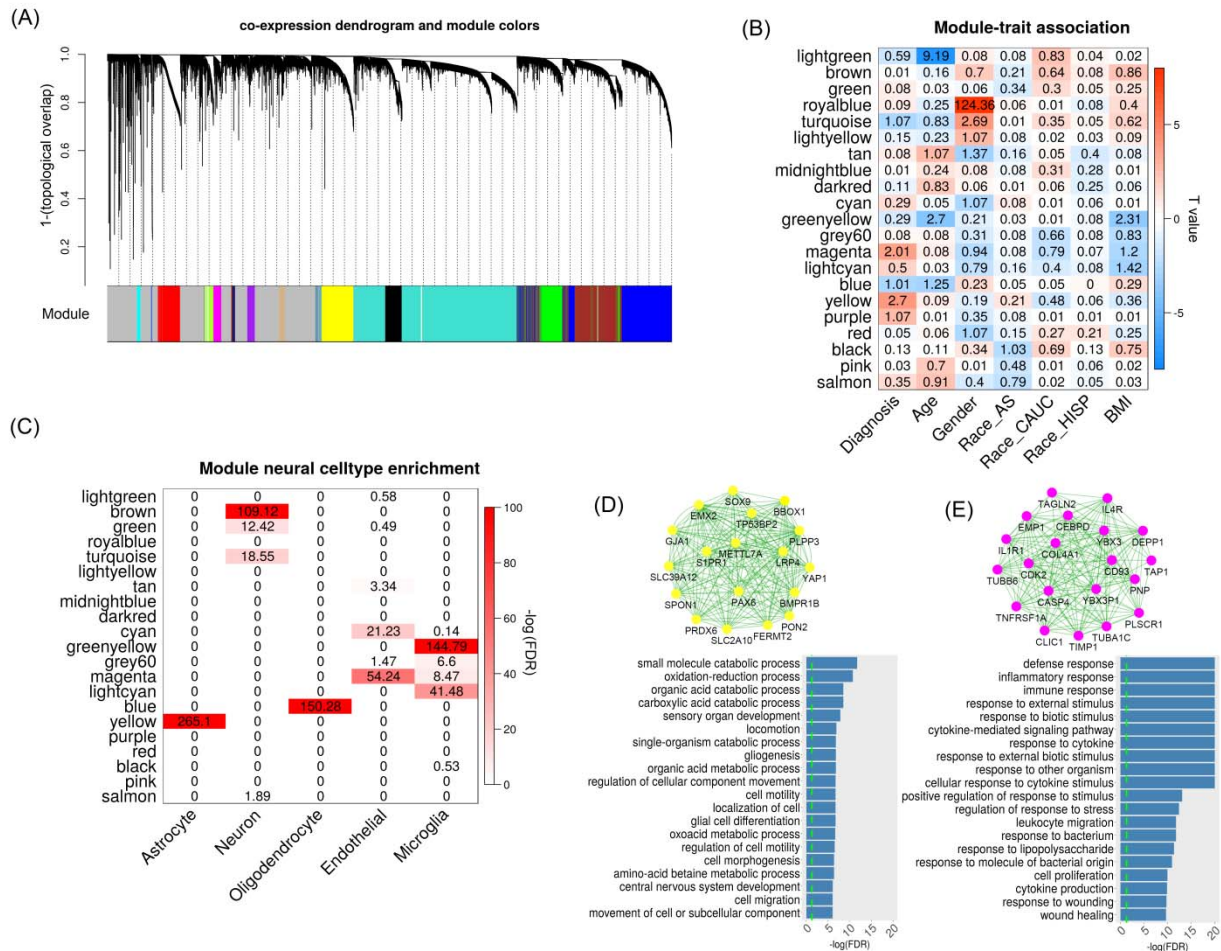


Figure.1 Co-expression analysis on brain DLPFC expression data. (A) shows the dendrogram clustering the correlated probes into several modules with different colors; (B) lists the relationship between module eigengenes and traits. The numbers indicate $\log_{10}(\text{FDR})$ and colors indicate the T-values. (C) shows the enrichment of module genes in seven reported modules which are specific to neural cells (i.e., neuron, endothelial, astrocyte, microglia and oligodendrocyte). (D) and (E) plot the network for the top 20 hub genes and their functional enrichment in gene ontology for magenta and yellow modules, respectively.

Methylation modules and their associations with SZ

By a similar network analysis, we identified 13 methylation modules and tested the associations of these module eigengenes with multiple traits as shown in Fig.2A. There were three modules colored by yellow (641 CpGs, $t\text{-stat} = -7.22$, $p = 6.6 \times 10^{-11}$), green (475 CpGs, $t\text{-stat} = 3.78$, $p = 7.5 \times 10^{-4}$) and turquoise (15983 CpGs, $t\text{-stat} = -2.87$, $p = 0.01$) showing significant relationship with schizophrenia diagnosis ($FDR < 0.05$). In each of three modules, CpGs tended to have more significant SCZ vs. HC group differences if they demonstrated higher membership within the module (correlations were 0.91, 0.61 and 0.43, respectively, Fig.S2 A-C). Fig.2C-D show the network structure of top 20 genes with high module membership in yellow and green modules. The genes from yellow module were significantly enriched in the functional pathways for developmental growth, axonogenesis, neurogenesis and neuron generations, while the genes from green module were related to cell projection, as shown in Fig.2C. Although genes in the turquoise module were also enriched in some neural development pathways such as forebrain development and central nervous system neuron differentiation, it had a large module size with less focus on specific functions and therefore was not further analyzed in this study. By matching the methylation nearby genes (<5kbp) with those from cell specifically expressed genes, we found strong enrichment of blue and red methylation modules in oligodendrocytes as shown in Fig.2B. Based on CpG patterns specific to neuron and glia cells, we identified that CpGs from blue and red methylation modules were more likely hypo-methylated in glia cells compared to neurons in human brain.

To validate the above results in *MethyRep1* data, we matched 95.48% of CpGs in yellow module and 97.47% of CpGs in green module. We did not find any significant group difference in the first eigengene of yellow module ($P = 0.28$) but we found that the second eigengene showed group differences ($t = -2.42$, $P = 0.02$). The disease association for green module was

significantly replicated with $t = 2.98$, $P = 7.1 \times 10^{-3}$. In addition, both modules were well preserved in the independent dataset with significantly higher GCC ($P < 1 \times 10^{-4}$, Fig.S3 E-F) and Z_summary statistics (yellow: $Z = 13$; green: $Z = 21$), as shown in Fig.S2 (D).

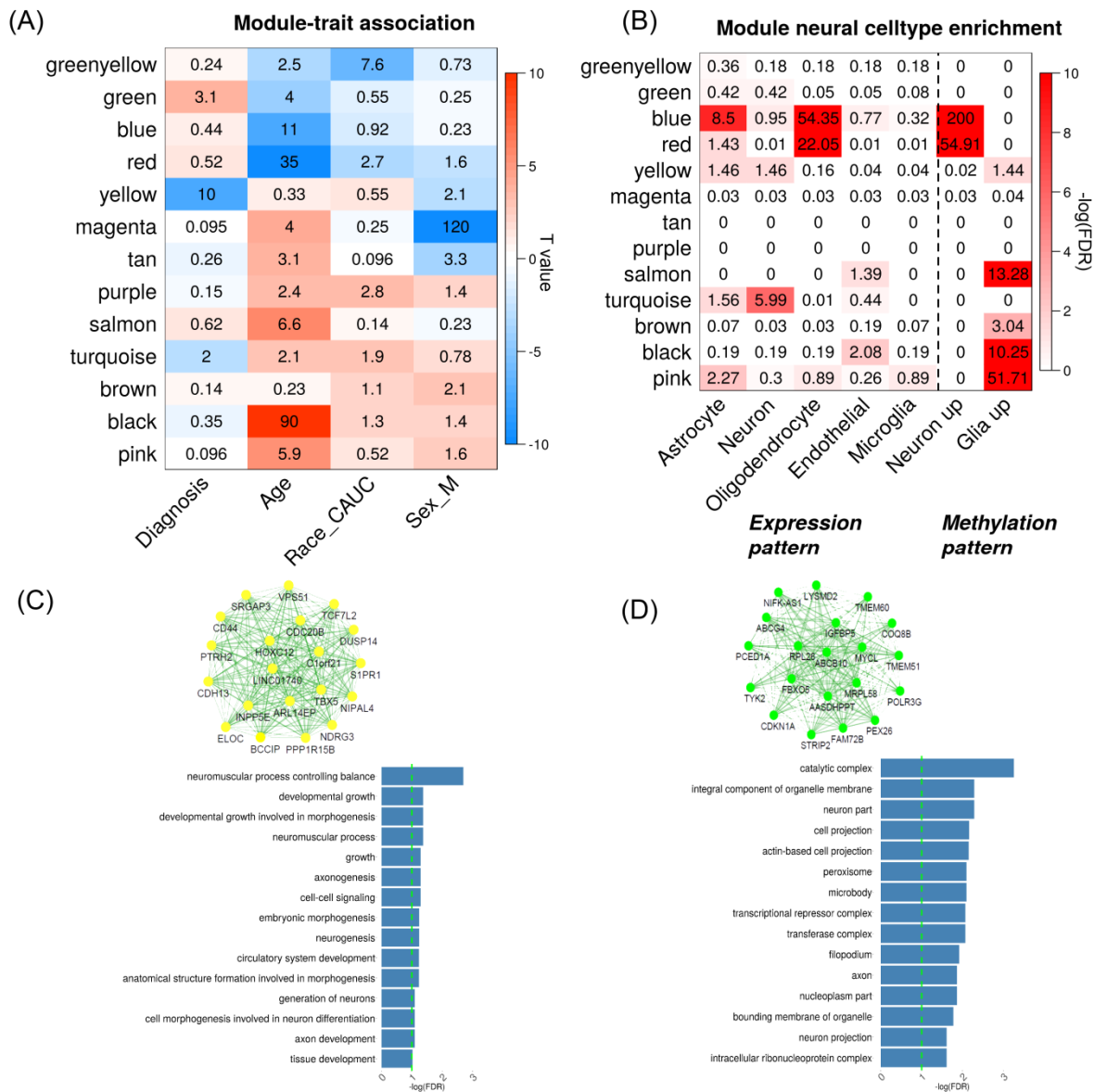


Figure 2. Network analysis on brain DLPFC methylation data. (A) shows the associations between module eigengenes and traits. The numbers indicate $\log_{10}(\text{FDR})$ and colors indicate the T-values. (B) lists the enrichment tests of each methylation module in different cell-specific

modules; and (C-D) plot the network of top 20 representative genes and their functional enrichment in gene ontology for yellow and green module respectively.

Overlap between expression and methylation modules

After annotating methylation sites and expression probes to their target gene ensemble IDs, we tested the overlap among the expression and methylation modules based on both Fisher's test (i.e. FDR_Fisher) and permutation test (i.e. FDR_perm). As shown in Fig.S5 (Supplementary file 1), we identified three sets of significantly overlapped modules which can be grouped to two categories: SCZ-related and age-related overlapping sets.

The SCZ-related overlapping set included the yellow expression module and yellow methylation module (odds ratio OR = 1.75, FDR_Fisher = 0.04, FDR_perm = 0.009). Both modules were significantly associated with SCZ disease as shown in Fig.1B and Fig.2A. There were 1109 expression probes (nearby 893 genes; Supplementary file 2) and 641 CpGs (nearby 591 genes; Supplementary file 3) from each module, respectively, with 5.2% probes located closely (<5kbp) to 9.5% CpGs, denoted by 'overlap CpGs' and 'overlap probes'. As shown in Fig.S6 (Supplementary file 1), overlap probes and CpGs were mainly involved the genes enriched in GO terms of regulating nitrogen compound metabolic process, biosynthetic process and fatty acid oxidation (FDR<0.05) and some interesting KEGG pathways like Wnt signaling, Adipocytokine signaling and Glutamatergic synapse(FDR<0.15). By annotating each CpG to its nearest gene, we found that 17.9% ($p = 1.3 \times 10^{-3}$) of CpGs were located within 100kbp of TSS of genes from the expression module, while 41.5% ($p = 2.2 \times 10^{-3}$) were within 500kbp, as shown in the density plot Fig.3A. Compared to those CpGs not included in the module, the overlap CpGs and the module CpGs(i.e., non-overlapped CpGs in the module) were significantly enriched in

CpG islands (CGI) north shore (odds ratio: OR = 1.38 and 1.72), south shore (OR = 1.5 and 1.36), TSS200 regions (OR = 1.17 and 1.24), and 5UTR (OR = 1.34 and 2.18) as shown in Fig.3B.

From previously reported eQTL and meQTL findings, we found that 32.8% of CpGs in the module (including 32% of overlap CpGs) were modified by cis-meQTLs, while 74% of expression probes (including 95.7% of overlap probes) were regulated by cis-eQTLs. And 12.6% of the cis-meQTLs are also cis-eQTLs for the expression module. In particular, 36.3% of cis-meQTLs modifying overlap CpGs, were also cis-eQTLs. In addition, 13 out of total 205 protein-coding genes in the methylation module were located in 108 SCZ risk regions [53] as listed in Table S1. The eQTLs also showed significant enrichment in risk loci of SCZ (OR = 1.31, $p_{\text{perm}} < 1 \times 10^{-5}$; $p_{\text{Fisher}} = 9.6 \times 10^{-11}$) and three other psychiatric disorders (Autism spectrum disorder (ASD), Attention-deficit/hyperactivity disorder (ADHD) and bipolar disorder (BIP)) as shown in Fig.3C. None significant enrichment was found for meQTLs in any psychiatric disorders.

Two age-related overlapping sets were found. One was between the blue expression module and the blue methylation module (OR = 2.39, $\text{FDR}_{\text{Fisher}} = 1 \times 10^{-39}$, $\text{FDR}_{\text{permutation}} < 1 \times 10^{-5}$), namely 'blueExp-blueMethy' link. The other was between the blue expression module and the red methylation module (OR = 5.79, $\text{FDR}_{\text{Fisher}} = 1 \times 10^{-30}$, $\text{FDR}_{\text{permutation}} < 1 \times 10^{-5}$), namely 'blueExp-redMethy' link. The blue expression module has a negative association with age as shown in Fig.1B (partial $\eta^2 = 0.025$, $p = 9.6 \times 10^{-3}$) and the association was replicated in *ExpRep1* (89.6% probes included, $p = 8.4 \times 10^{-3}$) and *ExpRep2* (61.2% genes included, $p = 4 \times 10^{-3}$). The blue and red methylation modules were also negatively significantly associated with age as shown in Fig.2A (blue: partial $\eta^2 = 0.19$, $p = 1.2 \times 10^{-11}$; red: partial $\eta^2 = 0.49$, $p = 2.5 \times 10^{-36}$) and

the age-methylation associations were also replicated (blue module: 96.6% CpGs included, $p = 0.016$; red module: 97% CpGs included, $p = 2.3 \times 10^{-6}$) in *MethyRep1*. GO term enrichment analysis showed that genes overlapped by blueExp-blueMethy modules were mainly enriched in functional pathways involving glial cell differentiation, gliogenesis, neurogenesis, axon ensheathment, and myelination (Fig.S7 (A)). Similar pathways were also significantly enriched by the overlapped genes from blueExp-redMethy linked modules, as shown in Fig.S7 (B).

In blueExp-blueMethy link, there are 2317 expression probes and 6308 CpGs with 20.9% CpGs located nearby 25.8% gene TSS (<5kbps). Fig.3D plots the distance distribution between CpGs and their nearest expression probes in the linked modules, showing that 36.4% CpGs were located within 100kbps of gene TSS and 65.5% CpGs were located within 500kbps. The overlap CpGs and module CpGs were mostly enriched in body of the gene and enhancer regions (OR = 1.77, $p = 8.1 \times 10^{-100}$ and OR = 1.85, $p = 1.4 \times 10^{-107}$, respectively) compared to all other CpGs, as shown in Fig.3E. In addition, 42.2% of CpGs from the module including 40.2% of overlap CpGs were significantly modified by cis-meQTLs, while 73.3% of expression probes including 94.9% of overlap probes were affected by cis-eQTLs. In particular, 46.2% of the cis-meQTLs are also cis-eQTLs for the expression module. Fig3.F shows the significance of eQTLs enriched in risk loci of SCZ, BIP, ADHD and ASD but none significant enrichment was found for meQTLs in any psychiatric disorders.

In the red methylation module (440 CpGs) which was also overlapped with blue expression module as shown in Fig.S8, 41.3% CpGs located closed to 5.4% of the expression gene TSS (<5kbps). 54.5% CpGs and 75.6% CpGs were located within 100kbps and 500kbps of TSS of genes from the expression module, respectively. CpGs from the red methylation module were significantly enriched in gene body (OR = 1.7, $p = 1.5 \times 10^{-7}$) and enhancer regions (OR =

2.8, $p = 2.9 \times 10^{-27}$) compared to all other CpGs. 54.5% of CpGs including 48.5% of overlap CpGs were modified by meQTLs, while 94.6% of overlap probes were affected by eQTLs. 35.7% of meQTLs were also cis-eQTLs in regulating the gene expression of the module.

Based on the *ExpRep1* and *MethyRep2* datasets assayed from the same subjects, we tested the associations between expression probes and their nearby (<500kpbs) CpGs from each of the above overlapping sets. There were only few expression-methylation pairs (1.2~8.9%) matched due to lower resolution in *MethyRep2*. After controlling for covariates age, sex and PMI, 20%, 25% and 47.8% of matched pairs showed significant associations ($FDR < 0.05$) in yellowExp-yellowMethy, blueExp-blueMethy and blueExp-redMethy links, respectively.

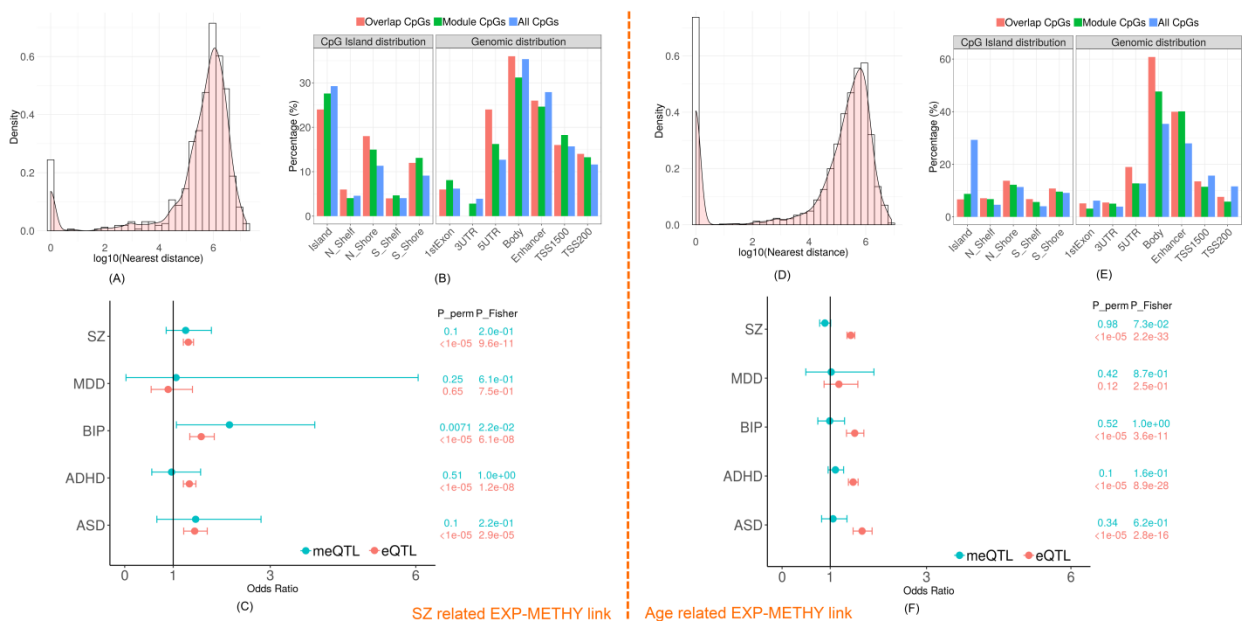


Figure 3. The characteristics of overlapping expression and methylation modules related to SCZ and age. (A) and (D) show the histogram plots of the distance between each methylation site and the nearest transcription start site of the linked modules. (B) and (E) plot the distribution of three sets of CpGs (overlap CpGs, module CpGs and all CpGs) across the genome. (C) and (F) list the

enrichment tests of cis-meQTL and cis-eQTL in risk loci of five psychiatric disorders (SZ: schizophrenia, MDD: major depressive disorders, BIP: bipolar disorder, ADHD: attention deficit hyperactivity disorder, ASD: autistic disorder) from PGC study.

Discussion

We used large transcriptomics and epigenetics data from human brain frontal cortex to explore the relationship between co-expression and methylation modules in terms of their overlap in genes, associations with SCZ and aging, as well as their shared underlying genetic risk. Our findings were replicated in multiple independent datasets on module's associations with SCZ and age, and module's preservation, and further verified by direct associations between individual CpGs and gene expression probes. We found two modules in each modality showing significant associations with SCZ, and three pairs of overlapping methylation and expression modules related to SCZ or aging.

SCZ-related modules and their expression-methylation relationship

Two co-expression modules were significantly associated with SCZ with one (yellow module) including the genes more specifically expressed in astrocytes. This module is highly preserved across studies and also in line with the identified modules in previous RNA-sequencing studies of DLPFC tissue [15, 16, 51], showing consistent pattern with up-regulation of module expression in SCZ compared to HCs. In particular, the genes from the yellow module are also highly enriched in the module (CD4: astrocyte) reported in a recent study of transcription across five psychiatric disorders [51], which demonstrated significant changes of the module expression in SCZ, autism, and bipolar disorder. Astrocytes are known to be involved in synaptic

metabolism and regulation of neurotransmitter release and reuptake (e.g., GABA and glutamate) and thereby critical for psychosis development[54]. Our pathway analysis found significant enrichment of the module genes in some SCZ-related KEGG metabolic pathways involving the metabolism on fatty acid [55], glycine [56], glutamate [57] and tryptophan [58], as shown in Fig.S9(A) (Supplementary file 1). Compared to the yellow module, the magenta module expression was also up-regulated in SCZ, but more specific to endothelial cells and microglia with genes mainly related with immune response. Previous studies have reported high involvement of immune system pathways in SCZ development via glia cells as well as their interactions with neurons in neurotransmitters perturbations [59]. The transcriptional changes of these immune system related genes suggest the complex co-expression patterns in downstream modulation of neuroinflammation in SCZ.

The yellow methylation module showed a significant association with SCZ. In particular, those methylation target genes included astrocyte highly-affinity glutamate transporters *SLC1A2* and *SLC1A3* [60], astrocyte-specific expressed marker gene *ALDH1L1* [61]. As shown in Fig.S9(B) and Table S2 (Supplementary file 1), they were also marginally enriched in GABAergic synapse (*GABRA1*, *GABRD*, *GNAO1*, *GNG7*, *PLCL1*, *PRKCA*, *CACNA1A*) and serotonergic synapse (*HTR1B*, *HTR2A*, *HTR3B*, *KCNJ3*), which are crucial neurotransmitter pathways in pathogenesis of SCZ [62-64]. Significant down-regulation of DNA methylation in most of CpGs near these genes (partial $R^2 = 1.8\% \sim 16.1\%$; Supplementary file 1, Table.S2) in SCZ, along with significant enrichment of module genes in growth development and neurogenesis, suggests the potential role of epigenetics in regulating glia cell function via these neurotransmitter signaling pathways. This is in agreement with previous findings of astrocyte epigenetic regulation in the pathophysiology of psychiatric disorders [54, 65, 66]. In addition, the

genes nearby CpGs in the module were more likely to be mapped to PGC SCZ risk regions ($p = 0.057$) than the genes from the other methylation modules. In particular for SCZ high risk genes (Supplementary file 1, Table S1) including *RERE*, *MAD1L1*, *FURIN*, *GATAD2A*, they have been reported to show significant methylation level alterations in SCZ from previous DNA methylation genome-wide association studies [67-69].

The significant overlap between the yellow expression and the yellow methylation modules mainly involves genes with functions in some metabolic processes, biosynthetic process and fatty acid oxidation, which demonstrated the potential relationship with nervous system development in previous studies [61, 70]. In particular, CpGs nearby genes *SLCIA2*, *SLCIA3* and *PRKCA* from glutamate synapse showed hypo-methylation in SCZ which may lead to up-regulated expressions of these genes in astrocyte of SCZ, consistent with our results as shown in Table.S1 (Supplementary file 1). Significantly, a higher percentage of CpGs in the module were enriched in CGI shores which is known to be variable, compared the CpGs from the other modules. In addition, cis-eQTLs were identified for a large proportion of genes and enriched for SCZ and other three psychiatric disorders risk loci. These results suggest combinatorial effects of genetic and epigenetic components in regulating gene expression in astrocytes, dysregulation of which is associated with SCZ.

Aging-related modules and their expression-methylation relationship

The module genes from blue expression module were highly enriched in genes expressed in oligodendrocytes, which are the CNS cells involved in myelination [71, 72]. GO enrichment test further validated that the functions of those specifically expressed module genes mainly involved oligodendrocytes differentiation, gliogenesis and development, ensheathment and myelination,

which have been reported to be critical in maintenance of white matter integrity and structural connectivity in central nervous system [73, 74]. We found down-regulation of module expression, especially of the oligodendrocyte *OLIG2* gene ($t = -2.56$, $p = 0.01$) along with increase of age. *OLIG2* can promote the formation of oligodendrocyte precursors and oligodendrocyte differentiation. Lower expression of *OLIG2* may cause deficits in oligodendrocyte production and differentiation and thus affect the formation of myelin, which is potentially related to the decreased ability of remyelination in the aging brain [75]. Although marginal association with SCZ was found in the first module eigengene, previous studies have shown the lower level of oligodendrocyte-associate gene expression in frontal cortex of SCZ along with myelination impairment and cognitive loss [6, 76]. Given that a number of studies have shown the dramatic loss of structural connectivity and white matter integrity across the brain regions in SCZ, it is promising to probe the role of oligodendrocytes-related expression module in the development of SCZ. In addition, the module harbored 13% SCZ risk genes and their cis-eQTLs were significantly enriched in most of psychiatric disorders risk loci, suggesting the potential pathology of genetic risk in SCZ through the alteration of module expression in oligodendrocytes.

Two methylation modules (blue and red) were significantly enriched in oligodendrocyte expressed genes and negatively associated with aging, pointing out the role of epigenetics in aging process by affecting oligodendrocyte development and differentiation. Previous work has reported the critical role of epigenetics in regulating oligodendrocyte gene expression and thereby leading to inefficient of remyelination during aging [77]. In our data DNA methylation of CpGs nearby myelin regulatory genes including *MYRF*, *MBP*, *MAG*, *MOG*, *CNP* [78], critical in oligodendrocyte development and myelination of axons, demonstrated significant negative

associations with age (partial $R^2 = 2\% \sim 18\%$; Supplementary file 1, Table S3) in the blue module. CpGs nearby genes *MYRF*, *MBP* and *MAG* also showed significant decreases (partial $R^2 = 9\% \sim 39\%$; Supplementary file Table S3) along with the increase of age in the red methylation module. Additional pathway analyses showed that those myelin related genes together with many other genes in the modules were enriched for several functional pathways including central nervous development, neurogenesis, and glia cell differentiation, suggesting the potential of epigenetic regulations of these pathways in altering the oligodendrocyte differentiation and myelinations during aging process [79].

Functional pathways related to oligodendrocyte development and differentiation were shared by both expression and methylation modules, leading to significant overlap between the modules. There was a significant proportion of CpGs (20.9~41.3%) located nearby TSS of expression module genes (<5kbp), mostly enriched in gene body and enhancer regions, suggesting the regulation effect of those methylation module in expression modules [29]. In particular, by the association tests on the methylation and gene expression from the same subjects, we identified 25~47.8% of matched expression-methylation pairs mainly showing significant negative associations (Supplementary file 1, Table S4), including some oligodendrocyte highly associated genes *MYRF*, *MAG* and *SOX10*. Enrichment tests on neural specific methylation patterns showed that both methylation modules were enriched for hypomethylation in glia cells compared to neurons, which might potentially contribute to their associated gene expression specificity in oligodendrocytes. These further evidences suggest epigenetic regulation on gene expression in oligodendrocytes to change myelination process. Genetics was also found to have large effects on both methylation and expression modules. A majority of module gene expressions were regulated by cis-eQTLs (74%~95%) and a relative

high proportion of CpGs (40.2~54.5%) from both methylation modules were significantly regulated by cis-meQTLs. Especially, 35.7~46.2% meQTLs modifying overlap CpGs were also cis-eQTLs, indicating a pathway of some genetics in regulating oligodendrocyte gene expression through epigenetics. This genetic, epigenetic and expression interplay is in line with our previous knowledge of both gene and environmental factors involved in the process of demyelination during aging [80]. In addition, we found significant enrichment of cis-eQTLs in psychiatric disorders' risk loci, which suggests shared genetics between risk for psychiatric disorders and regulating aging process which alter gene expression of oligodendrocytes, and partially reflect the similar decreases of myelination in both aging and SCZ patients.

To sum up, using a network analysis focused on DLPFC postmortem tissues, we identified and replicated two expression and two methylation modules significantly associated with SCZ with one pair showing significant overlap and enrichment in astrocyte-associated functional pathways. Additional overlapped expression-methylation module pairs were also replicated in relation to aging process with module genes mostly involved in oligodendrocyte development and myelination. Further examination of underlying QTLs showed the significant enrichment of eQTLs in most of psychiatric disorders' risk loci, but not for meQTLs, suggesting combinatorial effects of genetics and epigenetics in regulating gene expression in glia cells of psychiatric disorders. This work as proof-of-concept demonstrates the existence of coherence between gene expression and DNA methylation at the network level, and provides support for the critical role of epigenetics in glial cell development and function during aging process and psychosis development.

Limitations

The findings of this study should be interpreted in consideration of several limitations. Firstly, the analysis was performed on the datasets from different studies of different cohorts. Tests on the associations between methylation and expression from the same cohorts are still waiting for more efforts in psychiatric genetics research field. Although we tried to validate our findings in a study with both datasets, lower resolution of methylation array and missing of psychosis information limit our knowledge of the broad relationship between two features. Secondly, we have removed effects of some covariates (e.g., RNA quality, PMI, cell type and batch) prior network construction, but the modules may still be sensitive to the other factors (e.g., disease status and medication) and parameter settings in the method (e.g., block size and power index). Although we validated the preservation of identified modules in several independent datasets, more experiments are needed to evaluate the module robustness. Thirdly, genetic risk (eQTL and meQTL) were also discussed in the study, instead of computing eQTL from the data, we used the reports from public eQTL databases on RNAseq studies based on large sample size for eQTL detection. Although previous study has showed high overlap of RNAseq eQTLs with our microarray data [15], the results may still subject to the power and platform for eQTL detection.

References

1. Mathers C, Fat DM, Boerma JT: **The global burden of disease: 2004 update**: World Health Organization; 2008.
2. Sullivan PF, Kendler KS, Neale MC: **Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies**. *Archives of general psychiatry* 2003, **60**(12):1187-1192.
3. Gupta CN, Calhoun VD, Rachakonda S, Chen J, Patel V, Liu J, Segall J, Franke B, Zwiers MP, Arias-Vasquez A: **Patterns of gray matter abnormalities in schizophrenia based on an international mega-analysis**. *Schizophrenia bulletin* 2015, **41**(5):1133-1142.
4. Romme IA, de Reus MA, Ophoff RA, Kahn RS, van den Heuvel MP: **Connectome Disconnectivity and Cortical Gene Expression in Patients With Schizophrenia**. *Biol Psychiatry* 2017, **81**(6):495-502.
5. Calhoun VD, Eichele T, Pearlson G: **Functional brain networks in schizophrenia: a review**. *Frontiers in human neuroscience* 2009, **3**:17.
6. Flynn SW, Lang DJ, Mackay AL, Goghari V, Vavasour IM, Whittall KP, Smith GN, Arango V, Mann JJ, Dwork AJ *et al*: **Abnormalities of myelination in schizophrenia detected in vivo with MRI, and post-mortem with analysis of oligodendrocyte proteins**. *Mol Psychiatry* 2003, **8**(9):811-820.
7. Kelly S, Jahanshad N, Zalesky A, Kochunov P, Agartz I, Alloza C, Andreassen OA, Arango C, Banaj N, Bouix S *et al*: **Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group**. *Mol Psychiatry* 2018, **23**(5):1261-1269.
8. Keefe RS, Harvey PD: **Cognitive impairment in schizophrenia**. In: *Novel antischizophrenia treatments*. Springer; 2012: 11-37.
9. Bale TL, Baram TZ, Brown AS, Goldstein JM, Insel TR, McCarthy MM, Nemeroff CB, Reyes TM, Simerly RB, Susser ES *et al*: **Early life programming and neurodevelopmental disorders**. *Biol Psychiatry* 2010, **68**(4):314-319.
10. Horvath S, Janka Z, Mirnics K: **Analyzing schizophrenia by DNA microarrays**. *Biol Psychiatry* 2011, **69**(2):157-162.
11. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M: **A genome-wide association study of global gene expression**. *Nature genetics* 2007, **39**(10):1202.
12. Stranger BE, Raj T: **Genetics of human gene expression**. *Current opinion in genetics & development* 2013, **23**(6):627-634.
13. Ripke S, Neale BM, Corvin A, Walters JT, Farh K-H, Holmans PA, Lee P, Bulik-Sullivan B, Collier DA, Huang H: **Biological insights from 108 schizophrenia-associated genetic loci**. *Nature* 2014, **511**(7510):421.
14. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado-Torres L, Kam-Thong T, Xi HS, Quan J, Chen Q *et al*: **Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis**. *Nat Neurosci* 2018, **21**(8):1117-1125.
15. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, Ruderfer DM, Oh EC, Topol A, Shah HR *et al*: **Gene expression elucidates functional impact of polygenic risk for schizophrenia**. *Nat Neurosci* 2016, **19**(11):1442-1453.
16. Radulescu E, Jaffe AE, Straub RE, Chen Q, Shin JH, Hyde TM, Kleinman JE, Weinberger DR: **Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain**. *bioRxiv* 2018.
17. Katsel P, Davis KL, Gorman JM, Haroutunian V: **Variations in differential gene expression patterns across multiple brain regions in schizophrenia**. *Schizophr Res* 2005, **77**(2-3):241-252.

18. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, Schork AJ, Appadurai V, Buil A, Werge TM *et al*: **Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap.** *Science* 2018, **359**(6376):693-697.
19. Torkamani A, Dean B, Schork NJ, Thomas EA: **Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia.** *Genome research* 2010, **20**(4):403-412.
20. Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, Haeussler M, Sandoval-Espinosa C, Liu SJ, Velmeshev D *et al*: **Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex.** *Science* 2017, **358**(6368):1318-1323.
21. Aberg KA, McClay JL, Nerella S, Clark S, Kumar G, Chen W, Khachane AN, Xie L, Hudson A, Gao G *et al*: **Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults.** *JAMA psychiatry* 2014, **71**(3):255-264.
22. Vogel Ciernia A, LaSalle J: **The landscape of DNA methylation amid a perfect storm of autism aetiologies.** *Nature reviews Neuroscience* 2016, **17**(7):411-423.
23. Ramchandani S, Bhattacharya SK, Cervoni N, Szyf M: **DNA methylation is a reversible biological signal.** *Proceedings of the National Academy of Sciences* 1999, **96**(11):6107-6112.
24. Yao B, Christian KM, He C, Jin P, Ming GL, Song H: **Epigenetic mechanisms in neurogenesis.** *Nature reviews Neuroscience* 2016, **17**(9):537-549.
25. Delgado-Morales R, Agis-Balboa RC, Esteller M, Berdasco M: **Epigenetic mechanisms during ageing and neurogenesis as novel therapeutic avenues in human brain disorders.** *Clinical epigenetics* 2017, **9**:67.
26. Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, Kleinman JE: **Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex.** *Nat Neurosci* 2016, **19**(1):40-47.
27. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, Troakes C, Turecki G, O'donovan MC, Schalkwyk LC: **Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci.** *Nature neuroscience* 2016, **19**(1):48.
28. Andrews SV, Ellis SE, Bakulski KM, Sheppard B, Croen LA, Hertz-Picciotto I, Newschaffer CJ, Feinberg AP, Arking DE, Ladd-Acosta C *et al*: **Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder.** *Nat Commun* 2017, **8**(1):1011.
29. Lin D, Chen J, Perrone-Bizzozero N, Bustillo JR, Du Y, Calhoun VD, Liu J: **Characterization of cross-tissue genetic-epigenetic effects and their patterns in schizophrenia.** *Genome medicine* 2018, **10**(1):13.
30. Lipska BK, Deep-Soboslay A, Weickert CS, Hyde TM, Martin CE, Herman MM, Kleinman JE: **Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia.** *Biol Psychiatry* 2006, **60**(6):650-658.
31. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015, **43**(7):e47.
32. Hemani G, Shakhbazov K, Westra HJ, Esko T, Henders AK, McRae AF, Yang J, Gibson G, Martin NG, Metspalu A *et al*: **Detection and replication of epistasis influencing transcription in humans.** *Nature* 2014, **508**(7495):249-253.
33. Johnson WE, Li C, Rabinovic A: **Adjusting batch effects in microarray expression data using empirical Bayes methods.** *Biostatistics* 2007, **8**(1):118-127.
34. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD: **The sva package for removing batch effects and other unwanted variation in high-throughput experiments.** *Bioinformatics* 2012, **28**(6):882-883.

35. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: **Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.** *Bioinformatics* 2014, **30**(10):1363-1369.
36. Lin D, Chen J, Ehrlich S, Bustillo JR, Perrone-Bizzozero N, Walton E, Clark VP, Wang YP, Sui J, Du Y *et al*: **Cross-Tissue Exploration of Genetic and Epigenetic Effects on Brain Gray Matter in Schizophrenia.** *Schizophr Bull* 2017.
37. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R: **Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.** *Epigenetics* 2013, **8**(2):203-209.
38. Naeem H, Wong NC, Chatterton Z, Hong MK, Pedersen JS, Corcoran NM, Hovens CM, Macintyre G: **Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array.** *BMC Genomics* 2014, **15**:51.
39. Rahmani E, Yedidim R, Shenhav L, Schweiger R, Weissbrod O, Zaitlen N, Halperin E: **GLINT: a user-friendly toolset for the analysis of high-throughput DNA-methylation array data.** *Bioinformatics* 2017, **33**(12):1870-1872.
40. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9**:559.
41. Willer CJ, Li Y, Abecasis GR: **METAL: fast and efficient meta-analysis of genomewide association scans.** *Bioinformatics* 2010, **26**(17):2190-2191.
42. Durinck S, Spellman PT, Birney E, Huber W: **Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.** *Nat Protoc* 2009, **4**(8):1184-1191.
43. Hernandez DG, Nalls MA, Moore M, Chong S, Dillman A, Trabzuni D, Gibbs JR, Ryten M, Arepalli S, Weale ME *et al*: **Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain.** *Neurobiology of disease* 2012, **47**(1):20-28.
44. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J *et al*: **Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain.** *PLoS Genet* 2010, **6**(5):e1000952.
45. Narayan S, Tang B, Head SR, Gilmartin TJ, Sutcliffe JG, Dean B, Thomas EA: **Molecular profiles of schizophrenia in the CNS at different stages of illness.** *Brain research* 2008, **1239**:235-248.
46. McGirr A, Tousignant M, Routhier D, Pouliot L, Chawky N, Margolese HC, Turecki G: **Risk factors for completed suicide in schizophrenia and other chronic psychotic disorders: a case-control study.** *Schizophr Res* 2006, **84**(1):132-143.
47. Onnela JP, Saramaki J, Kertesz J, Kaski K: **Intensity and coherence of motifs in weighted complex networks.** *Physical review E, Statistical, nonlinear, and soft matter physics* 2005, **71**(6 Pt 2):065103.
48. Langfelder P, Luo R, Oldham MC, Horvath S: **Is my network module preserved and reproducible?** *PLoS computational biology* 2011, **7**(1):e1001057.
49. Dougherty JD, Schmidt EF, Nakajima M, Heintz N: **Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells.** *Nucleic Acids Res* 2010, **38**(13):4218-4230.
50. Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD, Vogel H, Steinberg GK, Edwards MS, Li G *et al*: **Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse.** *Neuron* 2016, **89**(1):37-53.
51. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, Schork AJ, Appadurai V, Buil A, Werge TM *et al*: **Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap.** *Science* 2018, **359**(6376):693-697.

52. Guintivano J, Aryee MJ, Kaminsky ZA: **A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression.** *Epigenetics* 2013, **8**(3):290-302.
53. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, Lin D-Y, Duan J, Ophoff RA, Andreassen OA: **Genome-wide association study identifies five new schizophrenia loci.** *Nature genetics* 2011, **43**(10):969.
54. Bernstein H-G, Steiner J, Guest PC, Dobrowolny H, Bogerts B: **Glial cells as key players in schizophrenia pathology: recent insights and concepts of therapy.** *Schizophrenia research* 2015, **161**(1):4-18.
55. Solberg D, Bentsen H, Refsum H, Andreassen O: **Association between serum lipids and membrane fatty acids and clinical characteristics in patients with schizophrenia.** *Acta Psychiatrica Scandinavica* 2015, **132**(4):293-300.
56. Sumiyoshi T, Anil AE, Jin D, Jayathilake K, Lee M, Meltzer HY: **Plasma glycine and serine levels in schizophrenia compared to normal controls and major depression: relation to negative symptoms.** *International Journal of Neuropsychopharmacology* 2004, **7**(1):1-8.
57. Marsman A, van den Heuvel MP, Klomp DW, Kahn RS, Luitjen PR, Hulshoff Pol HE: **Glutamate in schizophrenia: a focused review and meta-analysis of 1H-MRS studies.** *Schizophrenia bulletin* 2011, **39**(1):120-129.
58. Chiappelli J, Postolache TT, Kochunov P, Rowland LM, Wijtenburg SA, Shukla DK, Tagamets M, Du X, Savransky A, Lowry CA: **Tryptophan metabolism and white matter integrity in schizophrenia.** *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 2016, **41**(10):2587.
59. Khandaker GM, Cousins L, Deakin J, Lennox BR, Yolken R, Jones PB: **Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment.** *The Lancet Psychiatry* 2015, **2**(3):258-270.
60. Choudary PV, Molnar M, Evans SJ, Tomita H, Li JZ, Vawter MP, Myers RM, Bunney WE, Jr., Akil H, Watson SJ *et al*: **Altered cortical glutamatergic and GABAergic signal transmission with glial involvement in depression.** *Proc Natl Acad Sci U S A* 2005, **102**(43):15653-15658.
61. Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA: **A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function.** *Journal of Neuroscience* 2008, **28**(1):264-278.
62. Blum BP, Mann JJ: **The GABAergic system in schizophrenia.** *International Journal of Neuropsychopharmacology* 2002, **5**(2):159-179.
63. Ruzicka WB, Subburaju S, Benes FM: **Circuit- and Diagnosis-Specific DNA Methylation Changes at gamma-Aminobutyric Acid-Related Genes in Postmortem Human Hippocampus in Schizophrenia and Bipolar Disorder.** *JAMA psychiatry* 2015, **72**(6):541-551.
64. Aghajanian GK, Marek GJ: **Serotonin model of schizophrenia: emerging role of glutamate mechanisms.** *Brain research Brain research reviews* 2000, **31**(2-3):302-312.
65. Rajkowska G, Stockmeier CA: **Astrocyte pathology in major depressive disorder: insights from human postmortem brain tissue.** *Current drug targets* 2013, **14**(11):1225-1236.
66. Nagy C, Suderman M, Yang J, Szyf M, Mechawar N, Ernst C, Turecki G: **Astrocytic abnormalities and global DNA methylation patterns in depression and suicide.** *Mol Psychiatry* 2015, **20**(3):320-328.
67. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, 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**(1):176.

68. Montano C, Taub MA, Jaffe A, Briem E, Feinberg JJ, Trygvadottir R, Idrizi A, Runarsson A, Berndsen B, Gur RC *et al*: **Association of DNA Methylation Differences With Schizophrenia in an Epigenome-Wide Association Study.** *JAMA psychiatry* 2016, **73**(5):506-514.
69. So H-C: **Epigenome-wide association study and integrative analysis with the transcriptome based on GWAS summary statistics.** *arXiv preprint arXiv:170200329* 2017.
70. Maekawa M, Watanabe A, Iwayama Y, Kimura T, Hamazaki K, Balan S, Ohba H, Hisano Y, Nozaki Y, Ohnishi T: **Polyunsaturated fatty acid deficiency during neurodevelopment in mice models the prodromal state of schizophrenia through epigenetic changes in nuclear receptor genes.** *Translational psychiatry* 2017, **7**(9):e1229.
71. Gibson EM, Purger D, Mount CW, Goldstein AK, Lin GL, Wood LS, Inema I, Miller SE, Bieri G, Zuchero JB *et al*: **Neuronal activity promotes oligodendrogenesis and adaptive myelination in the mammalian brain.** *Science* 2014, **344**(6183):1252304.
72. Fields RD: **Oligodendrocytes changing the rules: action potentials in glia and oligodendrocytes controlling action potentials.** *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* 2008, **14**(6):540-543.
73. Hagmann P, Sporns O, Madan N, Cammoun L, Pienaar R, Wedeen VJ, Meuli R, Thiran JP, Grant PE: **White matter maturation reshapes structural connectivity in the late developing human brain.** *Proc Natl Acad Sci U S A* 2010, **107**(44):19067-19072.
74. Simons M, Nave K-A: **Oligodendrocytes: myelination and axonal support.** *Cold Spring Harbor perspectives in biology* 2016, **8**(1):a020479.
75. Sim FJ, Zhao C, Penderis J, Franklin RJ: **The age-related decrease in CNS remyelination efficiency is attributable to an impairment of both oligodendrocyte progenitor recruitment and differentiation.** *Journal of Neuroscience* 2002, **22**(7):2451-2459.
76. Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, Haroutunian V, Fienberg AA: **Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia.** *Proc Natl Acad Sci U S A* 2001, **98**(8):4746-4751.
77. Shen S, Sandoval J, Swiss VA, Li J, Dupree J, Franklin RJ, Casaccia-Bonnel P: **Age-dependent epigenetic control of differentiation inhibitors is critical for remyelination efficiency.** *Nature neuroscience* 2008, **11**(9):1024.
78. Emery B, Agalliu D, Cahoy JD, Watkins TA, Dugas JC, Mulinyawe SB, Ibrahim A, Ligon KL, Rowitch DH, Barres BA: **Myelin gene regulatory factor is a critical transcriptional regulator required for CNS myelination.** *Cell* 2009, **138**(1):172-185.
79. Liu J, Moyon S, Hernandez M, Casaccia P: **Epigenetic control of oligodendrocyte development: adding new players to old keepers.** *Current opinion in neurobiology* 2016, **39**:133-138.
80. Copray S, Huynh JL, Sher F, Casaccia-Bonnel P, Boddeke E: **Epigenetic mechanisms facilitating oligodendrocyte development, maturation, and aging.** *Glia* 2009, **57**(15):1579-1587.