

RESEARCH ARTICLE

Open Access



Transition of allele-specific DNA hydroxymethylation at regulatory loci is associated with phenotypic variation in monozygotic twins discordant for psychiatric disorders

Junping Ye^{1†}, Zhanwang Huang^{1†}, Qiyang Li^{1,2}, Zhongwei Li¹, Yuting Lan^{1,3}, Zhongju Wang¹, Chaoying Ni¹, Xiaohui Wu¹, Tingyun Jiang⁴, Yujing Li⁵, Qiong Yang³, Junghwa Lim⁵, Cun-Yan Ren⁶, Meijun Jiang⁷, Shufen Li¹, Peng Jin⁵, Jian-Huan Chen^{6*} and Cunyou Zhao^{1,2,7,8*} 

Abstract

Background Major psychiatric disorders such as schizophrenia (SCZ) and bipolar disorder (BPD) are complex genetic mental illnesses. Their non-Mendelian features, such as those observed in monozygotic twins discordant for SCZ or BPD, are likely complicated by environmental modifiers of genetic effects. 5-Hydroxymethylcytosine (5hmC) is an important epigenetic mark in gene regulation, and whether it is linked to genetic variants that contribute to non-Mendelian features remains largely unexplored.

Methods We combined the 5hmC-selective chemical labeling method (5hmC-seq) and whole-genome sequencing (WGS) analysis of peripheral blood DNA obtained from monozygotic (MZ) twins discordant for SCZ or BPD to identify allelic imbalances in hydroxymethylome maps, and examined association of allele-specific hydroxymethylation (AShM) transition with disease susceptibility based on Bayes factors (BF) derived from the Bayesian generalized additive linear mixed model. We then performed multi-omics integrative analysis to determine the molecular pathogenic basis of those AShM sites. We finally employed luciferase reporter, CRISPR/Cas9 technology, electrophoretic mobility shift assay (EMSA), chromatin immunoprecipitation (ChIP), PCR, FM4-64 imaging analysis, and RNA sequencing to validate the function of interested AShM sites in the human neuroblastoma SK-N-SH cells and human embryonic kidney 293T (HEK293T) cells.

Results We identified thousands of genetic variants associated with AShM imbalances that exhibited phenotypic variation-associated AShM changes at regulatory loci. These AShM marks showed plausible associations with SCZ or BPD based on their effects on interactions among transcription factors (TFs), DNA methylation levels, or other epigenomic

[†]Junping Ye and Zhanwang Huang contributed equally to this work.

Cunyou Zhao is the lead contact.

*Correspondence:

Jian-Huan Chen
cjh_bio@hotmail.com
Cunyou Zhao
cyzhao@smu.edu.cn

Full list of author information is available at the end of the article



marks and thus contributed to dysregulated gene expression, which ultimately increased disease susceptibility. We then validated that competitive binding of POU3F2 on the alternative allele at the ASHm site rs4558409 (G/T) in *PLLP*-enhanced *PLLP* expression, while the hydroxymethylated alternative allele, which alleviated the POU3F2 binding activity at the rs4558409 site, might be associated with the downregulated *PLLP* expression observed in BPD or SCZ. Moreover, disruption of rs4558409 promoted neural development and vesicle trafficking.

Conclusion Our study provides a powerful strategy for prioritizing regulatory risk variants and contributes to our understanding of the interplay between genetic and epigenetic factors in mediating SCZ or BPD susceptibility.

Keywords Schizophrenia, Bipolar disorder, 5hmC, Monozygotic twins, Allele-specific hydroxymethylation, *PLLP*

Background

Major psychiatric disorders, including schizophrenia (SCZ) and bipolar disorder (BPD), are complex genetic mental illnesses that affect more than 1% of people worldwide [1]. The complexity of these disorders arises from both their heterogenous symptoms and multifactorial: genetic, developmental, and environmental nature. Epidemiological and genetic studies have indicated substantial overlap and high relative risks among relatives of both SCZ and BPD patients [1]. Numerous genetic variants in noncoding regions implicated by genome-wide association studies (GWAS) are enriched in regulatory domains [2], indicating that regulatory role for nonsequence-based genomic variation in mediating associations among genetic risk burden, environmental or epigenetic risk exposure, and phenotype. Recently, the identification of nonimprinted allele-specific DNA methylation (ASM) associated with genetic variation in *cis*, where the genetic variant was associated with DNA methylation of a neighboring cytosine base on the same chromosome, suggested that mapping this type of allelic asymmetry may prove useful as a post-GWAS strategy for identifying functional genetic variants [3–5]. Our recent study based on monozygotic (MZ) twins discordant for SCZ or BPD revealed thousands of single-nucleotide polymorphisms (SNPs) identified with ASM imbalances and phenotypic variation-associated switching at regulatory loci and that affected the interaction among one or more transcription factors (TFs), DNA methylation levels, and likely other epigenomic mark levels [3]. Mapping of ASM or allele-specific chromatin states has also been reported to facilitate genome-wide screening for disease-linked regulatory SNPs [4, 6, 7], which can be prioritized for functional studies. Since there are two alleles on homologous chromosomes in an individual in the same cellular and nuclear environment, there might be heterogeneity between allelic epigenomic patterns among individuals, providing important implications for interindividual differences in disease susceptibility.

Recently, 5-hydroxymethylcytosine (5hmC), which is catalyzed from 5mC by ten eleven translocation (TET) proteins, was discovered to be a relatively abundant form

of cytosine modification in embryonic stem cells (ESCs) and Purkinje neurons [8, 9]. This discovery raised the possibility that 5hmC mark may function as an intermediate during DNA demethylation; in addition, it may be a regulatory epigenetic mark that alters chromatin structure or contributes to the recruitment or exclusion of other DNA-binding proteins that affect transcription, and therefore, the dysregulation of this mark may cause certain diseases, such as SCZ or BPD [10, 11]. Previously, sequence-dependent allelic imbalances in the epigenome, including imbalanced DNA methylation and histone marks or open chromatin at regulatory loci, have been reported to cause disease-associated switching, providing a powerful framework for identifying disease-associated variants and genes [3, 4, 7]. Although several groups have investigated the genomic distribution of DNA hydroxymethylation marks [12, 13], the role and functional importance of this modification and its links to disease-associated genetic variants in SCZ or BPD are still unclear.

We recently studied MZ twin epigenetic profiles and observed epigenetic variations, such as differences in DNA methylation [3], lncRNA [14], and miRNA [15] levels, in phenotypes among concordant or discordant MZ twins. Examination of the epigenetic profiles of MZ twins, particularly disease-discordant MZ twins, is a powerful strategy to gain understanding of how genetic, environmental, and stochastic factors impact epigenetic modifications and how epigenetic variations affect the acquisition of complex traits [3, 16]. MZ twins were exclusively matched for genotype, age, sex, paternal effects, population cohort effects, and exposure to several shared environmental factors. In summary, the same sources of cells from MZ cotwins were expected to exhibit identical genetic signatures, and alterations in allele-specific epigenetic modification in MZ cotwin pairs would indicate that certain epigenetic modifications may be particularly vulnerable to environmental influences or specific TFs, especially during embryonic development, leading to individual differences in phenotype and disease susceptibility [3, 14]. The effects of genetic variation on the stochasticity or metastability of the DNA

hydroxymethylome on the underlying heterogeneity of human disease have been unexplored.

In this study, we combined the 5hmC-selective chemical labeling method (5hmC sequencing [5hmC-seq]) [17] and whole-genome sequencing (WGS) of peripheral blood DNA obtained from MZ twins discordant for SCZ or BPD to identify allelic imbalances in hydroxymethylome maps and found mechanistic effects of allele-specific hydroxymethylation (AShM) transitions (gain or loss) at regulatory loci on the phenotypic variation of psychiatric disorders in discordant MZ twins (PDC twins). Determining the molecular pathogenic basis of these differences will contribute to our understanding of the interaction between genetic and epigenetic factors in mediating individual differences in disease susceptibility and provide a powerful strategy for prioritizing SCZ/BPD-associated genes or variants.

Methods

Participants

A total of 14 pairs of MZ twins (28 individuals), including six PDC twin pairs (three SCZ-discordant (SDC) twins and three BPD-discordant (BDC) twins), four psychiatric disorder-concordant (PCC) twins (one SCZ-concordant (SCC) twin and three BPD-concordant (BCC) twins), and four healthy concordant (HCC) twins, was recruited in this study (Additional file 1: Table S1). The zygosity was determined by the Qiagen Investigator Argus X-12 QS Kit (USA). All patients fulfilled the diagnostic criteria of SCZ or BPD according to the Diagnostic and Statistical Manual of Mental Disorder, 4th Edition (American Psychiatric Association). All participants in this study provided informed consent prior to the study following presentation of the nature of the procedures. Approval for the study was obtained from the local medical ethics committee and conducted in accordance with the Declaration of Helsinki.

WGS genotyping

To obtain genome-wide SNP genotype information, we performed WGS analysis of DNA from unaffected individuals of 14 MZ twin pairs with a depth of ~ 700 million 150-bp paired-end reads per sample on the Illumina Nova-seq by Novogene Solution (Tianjin, China). All sequencing data passed initial quality checks for base composition (no exclusions) using FASTQC and were mapped to hg19 using Bowtie2. We used Picard for removing duplicates and BaseRecalibrator and ApplyBQSR in GATK4 and public mutation datasets (1000G_phase1, Mills_and_1000G_gold_standard.indel, dbsnp_146) for validation of BAM files to obtain base quality scores. We used HaplotypeCaller in GATK4 for genotype determination and AlternativeRecalibrator and

ApplyVQSR for cross-validation to obtain the final genotype of each SNP.

Genome-wide analysis of AShM

We performed genome-wide DNA hydroxymethylation analysis using 5hmC-selective chemical labeling method (5hmC sequencing [5hmC-seq]) as previously described [12, 17]. Enriched DNA from 5hmC capture were subjected to library construction and to sequencing to a depth of ~ 26 million 51-bp single-end reads per sample on an illumine Hiseq by Annoroad Gene Technology (Beijing, China). All sequencing data passed initial quality checks for base composition (no exclusions) using FASTQC and was mapped to hg19 using Bowtie2 [18]. After removing duplicates using Picard, we quantified hydroxymethylation levels using MEDIPS [19] to produce the short read counts and the mean relative hydroxymethylation score (rhms) in each 500-bp bin (overlap of 250 bp) across the genome. The VarScan [20] package was used to detect the coverage, quality, and allele frequency of alternative allele from unique mapped data of 5hmC-seq followed by WGS genotyping. A genomic locus would be called out as a candidate polymorphic site if there was at least one high-quality read supporting the alternative allele. To reduce false positives, we only focused on candidate alternatives that had been called in WGS genotyping and showed concordant genotype in 28 MZ twins. We applied the Bayesian generalized additive linear mixed model approach (implemented using the INLA package) [21, 22] to quantify AShM based on 5hmC-seq read counts as described in our previous study [3]. For the individual AShM analyses, data were subset by AShM, and a binomial mixed model was fit with counts of reference and alternative reads as the outcome variable and random effects of disease status (β_s) and twin subject (γ_i): $\text{logit}(p_{is}) = \beta_0 + \beta_s + \gamma_i$, where p_{is} is the proportion of reads supporting the alternative allele in twin i and disease status s . The posterior probability distribution of the intercept term (β_0) was used to quantify the degree of AShM. Credible intervals were estimated by taking empirical quantiles of these distributions, and two-tailed posterior predictive p values were calculated based on the areas in the tail more extreme than the null alternative allele ratio of 0.5. Based on these p values, we used the Benjamini–Hochberg procedure to control the FDR at 10%. The SNP sites with FDR less than 0.1 were considered as the allele-specific hydroxymethylation bias sites (AShM sites). To validate the individual AShM sites that showed disease-associated AShM switching, we tested for each AShM whether a model including disease status as a fixed effect (M_1 : $\text{logit}(p_{is}) = \beta_0 + \beta_s + \gamma_i$) fit the data significantly better than a reduced model without this term (M_0 : $\text{logit}(p_{is}) = \beta_0 + \gamma_i$) among six PDC

twins, where p_{is} is the proportion of reads supporting the alternative allele in twin i and disease status s and γ_i is the random effect of twin. We then compared the likelihoods of these two models to generate a Bayes factor (BF):

$$BF_{1,0} = \frac{p(\text{data}|\text{M1})}{p(\text{data}|\text{M0})}$$

We considered AShM sites with $BF > 1$ as discordant AShM sites between affected and unaffected individuals, AShM sites with $BF > 10$ were considered as discordant AShM sites with strong evidence of disease-associated AShM transition.

Functional enrichment analysis of psyAShM sites

We used ANNOVAR software to annotate those alternatives called by Varscan and performed filter-based annotation and gene-based annotation to identify whether an alternative is reported in dbSNP and genomic context for each alternative. Gene Ontology term and pathway enrichment analysis was performed using the functional enrichment tool at <https://toppgene.cchmc.org/enrichment.jsp> [23].

To test enrichment of AShM sites among specific chromatin states, we obtained ten human brain tissues' chromatin state from the chromHMM dataset using the 15-state model [24, 25]. The 15-state model consists of eight active states associated with gene expression (active TSS, TssA; flanking active TSS, TssAFlnk; transcription at 5' and 3' ends of genes, TxFlnk; strong transcription, Tx; weak transcription, TxWk; genic enhancers, EnhG; enhancers, Enh; Zinc finger genes/repeats, ZNF/Rpts) and seven repressed states (heterochromatin, Het; bivalent/poised TSS, TssBiv; flanking bivalent TSS/enhancer, BivFlnk; bivalent enhancer, EnhBiv; repressed Polycomb region, ReprPC; repressed PolyComb region ReprPCWk; quiescent, Quies). We calculated the number of SNPs located within each of 15 chromosome states for 807 psyAShM sites, and then enrichment was performed by testing whether the proportion of 807 psyAShM sites (Fisher's exact test) within chromosome states were significantly higher compared with 117,012 AShM sites, and meta-analysis was performed to combine chromatin state enrichment results from different brain tissue enrichment results. Error bar plots were generated for visualization of meta results.

We then employed DeepSEA [26] to predict AShM sites occurrence within regulatory enhancer or promoter regions. An "e-value" is calculated by DeepSEA based on the empirical distribution of that feature's effect ($\text{abs}(\text{palt} - \text{pref})$) among gnomAD alternatives to evaluate the expected proportion of SNPs with a larger predicted effect. We set an e-value of 0.01 as a significant effect, which indicates that only 1% of gnomAD alternatives have a larger predicted effect as a threshold to obtain a significant effect of SNPs on chromatin states. Z-score is

a scaled score where the feature diff score ($\text{palt} - \text{pref}$) is divided by the root mean square of the feature diff score across gnomAD alternatives. Note that this is "sign-preserving"; i.e., a negative z-score indicates that a mutation decreases the probability of a regulatory feature. We focused on the z-scores of H3K27me3, H3K9me3, H3K4me3, H3K4me1, H3K27ac, and H3K36me3 histone markers in the human hippocampus, female fetal brain, and male fetal brain from DeepSEA [26], filtered alternatives outside the enhancer or promoters, and calculated the number of SNPs with the same or opposite direction as in 807 psyAShM sites. In detail, if SNPs whose alternative alleles have both positive/negative effects on enriched histone modifications of the above six histone markers while alternative allele frequency more/less than 0.5, the SNPs would be defined as having the same effect on histone modifications. Otherwise, the SNPs would be defined as having the opposite effect on histone modifications. Binom test is further used to test the consistency of effect direction for alternative allele on histone modification and DNA hydroxymethylation based on the binomial distribution.

We employed motifbreakR [27] to determine whether 807 psyAShM sites significantly overlapped a sequence match against a motif library of 431 human TF previously generated via the high-throughput systematic evolution of ligands by exponential enrichment method [28]. ScoreRef and scoreAlt were scores determined by the scoring method, when the sequence contains the reference/alternative SNP allele. The scores are scaled as a fraction of scoring range 0–1 of the motif matrix. We used feature's effect difference ($\text{abs}(\text{scoreRef} - \text{scoreAlt})$) among alleles to evaluate the effect of SNP on TF, then we analyzed the correlations between feature's effect difference and hydroxymethylation levels of alternative allele with Pearson's correlation (TF was considered significant correlation at $p < 0.05$).

Four kinds of dataset, including SCZ+BPD, Alzheimer disease (AD), BPD, major depressive disorder (MDD), from transcriptome-wide association study (TWAS)-hub at <http://twas-hub.org/> were used for enrichment analysis. The genes in these five gene sets were reported as potentially high risk genes related to SCZ and BPD, including activity-regulated cytoskeleton-associated protein (ARC), the N-methyl-d-aspartate receptor complex (NMDAR), fragile X mental retardation protein (FMRP) targets, postsynaptic density (PSD) proteins, and gamma-aminobutyric acid (GABA) [29–32]. Differential expressed genes were obtained from PsychENCODE consortium [33], CommonMind Consortium (CMC) [34], and lymphoblastoid cell lines (LCLs) [35]. We downloaded GTEx (Genotype-Tissue Expression) v8 eQTL dataset including 10 brain tissues, LIBD

(Lieber institute for brain development) eQTL dataset including hippocampus and dorsolateral prefrontal cortex (DLPFC), and human brain mQTL dataset. GWAS datasets include PGC3 SCZ [36], PGC3 BPD [37], attention deficit hyperactivity disorder (ADHD) [38], and autism spectrum disorder (ASD) [39], and MDD [40]. We calculated the overlapping gene numbers with psyAShM genes and performed enrichment analysis with Fisher's exact test using 150,000 SNP associated genes selecting by random sampling as background.

Cell culture and luciferase reporter assay

The human SK-N-SH neuroblastoma cells, the most widely used system to study neurodevelopment [3, 41–45], have the ability to differentiate into a neuronal phenotype characterized by extensive neurite outgrowth, and were used to investigate the function of psyAShM sites in early neurodevelopmental processes and endocytic membrane trafficking. The cells were cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS; FSP500, ExCell Bio) and maintained at 37°C with 5% CO₂. For differentiation, SK-N-SH cells was induced with normal DMEM supplemented with 1.25% FBS and 10 μM all-trans retinoic acid (ATRA, Sigma) for 3 days after the cells had been seeded into a T175 flask in normal DMEM with 10% FBS, and the differentiation medium was changed every 24 h. The cells were photographed with an inverted fluorescent microscope (Nikon Ti-U, ×20). The neurite-like length was measured manually in a blinded manner with ImageJ software (<https://imagej.nih.gov/ij/>). We used the tool of "straight" to depict the neurite on SK-N-SH from the cell body to the end of neurite and the software would calculate the length through the bottom of analysis-measure. To examine the regulatory role of psyAShMs, we cloned DNA fragments spanning ~1000 bp upstream to ~600 bp downstream of psyAShM site identified in the enhancer region of a pGL4.23 luciferase reporter containing the minimal TATA-box promoter (minP). Mutations at the psyAShM site were achieved by using the site-directed mutagenesis method by replacing a major allele with a minor allele at the given SNP sites (Additional file 1: Table S2). These reporter constructs were transiently cotransfected into SK-N-SH or HEK293T cells together with the pRL-TK plasmid as an internal control for transfection efficiency using Lipofectamine 2000 reagents (11668019, Thermo Fisher). Cells were harvested 48 h after transfection, and the dual luciferase activity (Promega, E1960) was measured with the Victor X multilabel readers (PerkinElmer, San Jose, CA, USA).

Electrophoretic mobility shift assay (EMSA), chromatin immunoprecipitation (ChIP), and PCR

Nuclear protein extraction from HEK293T cells was performed with EMSA to evaluate the affinity of allelic variations or hydroxymethylation alteration of rs4558409 (G/T) loci by using the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Fisher, USA). The biotin-labeled DNA complexes were visualized and quantified by using a chemiluminescence imaging system (Tanon 5200). ChIP assays were performed with cell extracts from HEK293T using anti-POU3F2 antibodies (1:50, Cell Signaling Technology, #12137). Cell lysates extracted from HEK293T cells were used with 2–4 μg antibodies for each assay and incubated overnight at 4°C. Protein G/Protein A agarose beads (Invitrogen™) were added for 3 h and quantified by qPCR (Additional file 1: Table S2).

Lentivirus plasmid construction, packaging, and transduction

To examine the role of SNP rs4558409 in *PLLP*, we designed two sgRNA to disrupt this site from the genome based on CRISPR/Cas9 technology using online tools from the Zhang laboratory (<https://zlab.bio/guide-design-resources>). After synthesis from Invitrogen, the sgRNA oligonucleotides were cloned into a transfer lentiviral vector (lentiCRISPRv2, Addgene #52961) via a BsmBI site (NEB, SN: R0739). To package lentivirus, we cotransfected HEK293T cell with three plasmids (pSPAX2, Addgene#12259; pMD2G, Plasmid #12259 and lentiCRISPRv2 which carried sgRNA for *PLLP*) by polyethylenimine (YEASEN, SN: 40816ES02, 1 mg/ml). After 60 h transfection, the cell culture medium was collected and centrifuged at 10,000 rpm for 10 min, and the supernatant were further filtered with a 0.22-μm filter through the ultracentrifugal tubes (Crystalgen, SN: 23-2590) at 20,000 rpm for 2 h (Beckman Avanti J-E, rotor JA-20). The sediment was resuspended by 20 μl PBS and stored in –80°C. The SK-N-SH cells in a 12-well plate were infected with lentivirus. The cells were incubated for an additional 48–72 h prior to identification of GFP+ cells under an inverted fluorescence microscope, followed by screening for stable cell lines with 1 μg/ml puromycin (biosharp, SN: BS111-25mg). The effects of rs4558409 disruption on *PLLP* expression were confirmed by qRT-PCR and Western blotting with rabbit anti-*PLLP* (1:500, LifeSpan BioSciences LS-C398159) and mouse anti-GAPDH (1:50,000, Proteintech, 60004-1-Ig) as a loading control.

FM4-64 imaging of SK-N-SH cells

We performed the experiments of endocytic membrane trafficking in SK-N-SH cells by FM4-64 unloading as

described previously [46]. Culture medium of SK-N-SH cells were replaced with saline solution containing 170 mM NaCl, 3.5 mM KCl, 0.4 mM KH_2PO_4 , 5 mM NaHCO_3 , 1.2 mM Na_2SO_4 , 1.2 mM MgCl_2 , 1.3 mM CaCl_2 , 5 mM glucose, and 20 mM N-tris(hydroxymethyl)-methyl-2-aminoethane-sulfonic acid (pH 7.4) and incubated for 10 min. After that, the cells were loaded with 10 μM FM4-64 (Invitrogen) for 2 min in saline solution supplemented with 75 mM KCl. The cells were rinsed with saline solution only and then incubated with 10 μM FM4-64 in saline solution for 10 min. The cells were perfused three times with saline solution for a total of 5 min and then left in saline solution for an additional 10 min. FM4-64 imaging was performed on an LSM 880 with an Airyscan confocal microscope (Zeiss) with a C-Apochromat $\times 40/1.2$ W Korr FCS M27 objective with images taken every 1.26 s at 25°C. A1-min baseline was recorded, followed by stimulation with 75 mM KCl in saline solution for 8 min. Cells were excited at 488 nm, and the emission was measured at 562 nm. The FM4-64 signal was determined by $F = (F_1/(F_0 - B_0))$. The signal was normalized to the mean fluorescence intensity measured at baseline.

RNA-Seq analysis

RNA extracted from rs4558409-KO or wild-type SK-N-SH cell lines was employed for RNA-seq analysis on the Illumina Nova-PE150 by Novogene Solution (Tianjin, China) with a depth of ~ 30 million 150-bp paired-end reads per sample. The raw reads were subjected to quality control with FASTQC and then clean reads was aligned to the hg19 reference genome using HISAT2 with default parameters and converted the sam file to the bam file (binary format) using samtools. Read counts were derived from the GTF file generated by Stringtie with the parameters -A, -e and using the Python scripts provided by Stringtie. Differentially expressed genes (DEGs) were generated by DESeq2 in R. Gene Ontology (GO) analysis of significant genes were enriched using an online database (ToppGene) and visualized using the ggplot2 package in R.

Results

AShM sites showed SCZ/BPD-associated transitions in MZ twin pairs

To identify individual SNP related to psychiatric disorder-associated ASHM transitions, we performed 5hmC-seq (at a depth of ~ 26 million 51-bp single-end reads per sample) and WGS (at a depth of ~ 700 million 150-bp pair-end reads per sample; Fig. 1) with peripheral blood DNA obtained from 14 MZ twin pairs (28 individuals. Additional file 1: Table S1). We then performed a genome-wide DNA hydroxymethylation pattern evaluation with samples from the MZ twins, and observed that

within-twin patterns of DNA hydroxymethylation levels were highly correlated among all MZ twin pairs (average r across all 500-bp windows within twins = 0.87; Additional file 2: Fig. S1), indicating high consistency in genome-wide DNA hydroxymethylation. We further evaluated the evidence for the sensitivity of the hydroxymethylome to genetic variation (viz. ASHM) in 14 MZ twin pairs and discovered 9,045,277 informative SNPs with heterozygous genotypes in at least one twin pair, as determined by WGS genotyping. We used these data for further analysis of ASHM on the alternative allele of each SNP (Fig. 1A and Additional file 1: Table S3). For each of these informative SNPs, we calculated a proportion of reads supporting evidence for an alternative allele in heterozygous individuals based on 5hmC-seq data via VarScan and then applied a Bayesian approach to assess DNA hydroxymethylation patterns exhibiting asymmetry between two alleles in each heterozygous individual (Fig. 1B). We discovered that 117,012 of 9,045,277 informative SNPs exhibited ASHM patterns with a false discovery rate (FDR) criterion of 10% in at least one of 28 individuals; these SNPs included 53,425 ASHM SNPs in one of the two individuals in at least one PDC twin pair, 50,967 SNPs in at least one PCC twin pair, and 45,068 SNPs in at least one HCC twin pair (Additional file 1: Table S3). Of the 117,012 identified ASHM SNPs, 3639 SNPs exhibited concordant ASHM patterns, and 16,282 SNPs exhibited discordant ASHM patterns in two individuals in at least two twin pairs (Fig. 1A and Additional file 1: Table S3). Of the 117,012 identified ASHM SNPs, 13,649 SNPs showed an ASM pattern ($\text{OR}=5.84, P < 2.2e-16$) as indicated in our recent report using data obtained from the same MZ twin cohort [3], and 534 ASHM SNPs were located at 53 previously reported imprinted loci (Additional file 1: Table S4; <https://www.geneimprint.org/>). Moreover, eight ASHM SNPs had exhibited allele-specific open chromatin (ASoC) in a previous assay for transposase-accessible chromatin using sequencing (ATAC-seq) [7]. These results support the idea that an allelic imbalance in the epigenome is a common phenomenon in the human genome and that the sensitivity to genetic variation or environmental influences that this epigenome imbalance confers may play roles in human diseases.

To examine whether ASHM was associated with phenotype-associated transitions in the two individuals in a PDC twin pair, we applied Bayes factors (BF) derived from the Bayesian generalized additive linear mixed model [21, 22] to 8544 ASHM SNPs to identify group-level ASHM transition patterns among PDC twin pairs (Fig. 1B). We identified 807 SNPs ($\text{BF}>1$) and 200 SNPs ($\text{BF}>10$, Fig. 1C) that underwent a significant ASHM transition from unaffected to affected

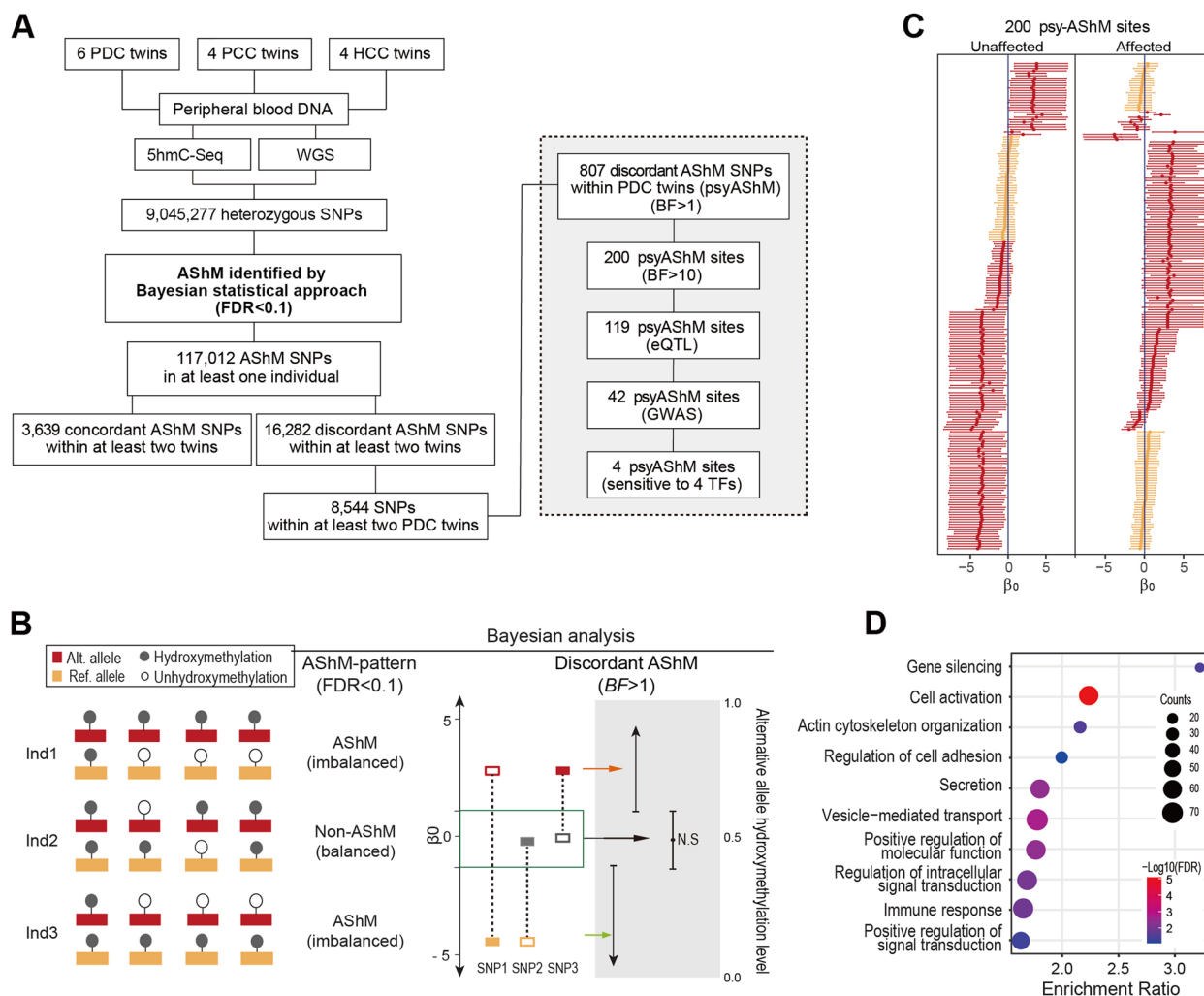


Fig 1 Schematic describing the method of detecting AShM in MZ twin pairs. **A** A schematic of the pipeline for the analysis of AShM. **B** Calling allelic imbalances. Allelic DNA hydroxymethylation patterns are shown in the left panel. Discordant AShM sites in twin pairs were determined by Bayesian analysis. **C** Estimates of 200 discordant AShM sites in PDC twin pairs. Each point represents the β_0 value of the psyAShM sites showing the greatest evidence of AShM in unaffected or affected individuals (highlighted in red, $p < 1e-10$) and the discordant AShM pattern within PDC twin pairs (BF>10). Error bars indicate 95% credible intervals. **D** Functional enrichment of GO-Biological process (GO-BP) annotations of 807 psyAShM sites

individuals among the PDC twin pairs (viz. psychiatric disorder-associated AShM sites, psyAShM sites; Additional file 1: Table S5). Of the 807 identified psyAShM sites, 374 showed AShM on/off switching (199 with a gain of AShM and 175 with a loss of AShM), 177 SNPs showed AShM flipping from unaffected individuals to the affected individual among the PDC twins, and 256 SNPs showed similar AShM patterns between unaffected and affected individuals among the PDC twins. To describe the common features of the 807 psyAShM sites annotated in 484 unique genes, we performed Gene Ontology-Biological process (GO-BP)

enrichment analysis with the 484 genes and identified a total of 41 GO-BP terms that passed a BH-adjusted P value<0.05 threshold of significance (Additional file 1: Table S6). The gene sets with the most enriched terms were involved in gene silencing, cell activation, actin cytoskeleton, regulation of cell adhesion, secretion, vesicle-mediated transport, positive regulation of molecular function, regulation of intracellular signal transduction, immune response, and positive regulation of signal transduction (Fig. 1D), part of which have been reported to be associated with SCZ/BPD [14, 46], providing functional implications for the psyASM sites in the development of these disorders.

The psyAShM transition displays SCZ/BPD-associated features

We then determined the genomic features of 807 psyAShM sites that are characteristic of SCZ or BPD risk genes. First, we observed that 807 psyAShM sites exhibited significant enrichment in lymphoblastoid cell lines (LCLs)-derived SCZ-associated DEGs (354 overlapping genes, odds ratio (OR) =1.84 and $P=4.3e-17$; Fig. 2A) [35], significant enrichment in brain-derived SCZ-associated DEGs (225 overlapping genes, OR=1.3 and $P=0.0021$), and marginal enrichment in brain-derived BPD-associated DEGs (73 overlapping genes, OR=1.3 and $P=0.05$) based on the PsychENCODE brain RNA-seq dataset [47]. Those brain-derived DEGs containing psyAShM sites, such as *PDE4B*, *RGS12*, *STXBP2*, and *SERPINE2*, are associated with cellular components of neuron projection, dendrite, synapse, postsynaptic density, and secretory vesicle, indicating the involvement of psyAShM transitions in the dysregulation of neuron-related gene expression in SCZ or BPD. We further assessed the enrichment of these psyAShM loci in two gene sets showing significantly enhanced ORs, including FMRP targets ($P=1.7e-4$, OR = 1.8) and PSD proteins ($P=2.7e-3$ and OR = 1.5; Fig. 2B), which have been previously implicated in SCZ or BPD [30, 31]. We also observed 807 psyAShM loci enriched significantly in a TWAS of SCZ+BPD patients ($P=1.8e-3$, OR =

1.4; Fig. 2C) [48]. These psyAShM sites were ultimately observed to be significantly enriched in subthreshold GWAS SNPs ($P<0.05$, Fig. 2D) of MDD ($P=4.9e-28$, OR=5.1), SCZ ($P=6.1e-8$, OR=1.8), and BPD ($P=2.5e-8$, OR=2.0) PGC GWAS loci [36–40]. Altogether, these results suggest that psyAShM loci are associated with psychiatric disorders and emphasize the value of using AShM data for prioritizing risk genes or variants.

psyAShM transitions exerts epiallele-specific effects on chromatin states and TF binding affinities

To gain insight into the global patterns of psyAShM sites, we examined their distribution using ANNOVAR software, which was used to annotate the specific genomic context of the 807 psyAShM sites (Fig. 3A). We found that the majority of the identified 807 psyAShM sites were located in noncoding regions, including intronic regions (51.67%), intergenic regions (33.46%), and non-coding RNA intronic regions (4.96%). Intriguingly, 807 psyAShM sites were also significantly enriched in eQTLs and brain mQTLs, and in genes expressed specifically in the brain or blood (Fig. 3B), further supporting their potential regulatory roles. We then observed that these psyAShM sites showed a bias toward alternative allele hydroxymethylation (alt/ref=2.07 for 8544 AShM sites and 2.87 for 807 AShM sites), and this preferential pattern of alternative allele hydroxymethylation showed a

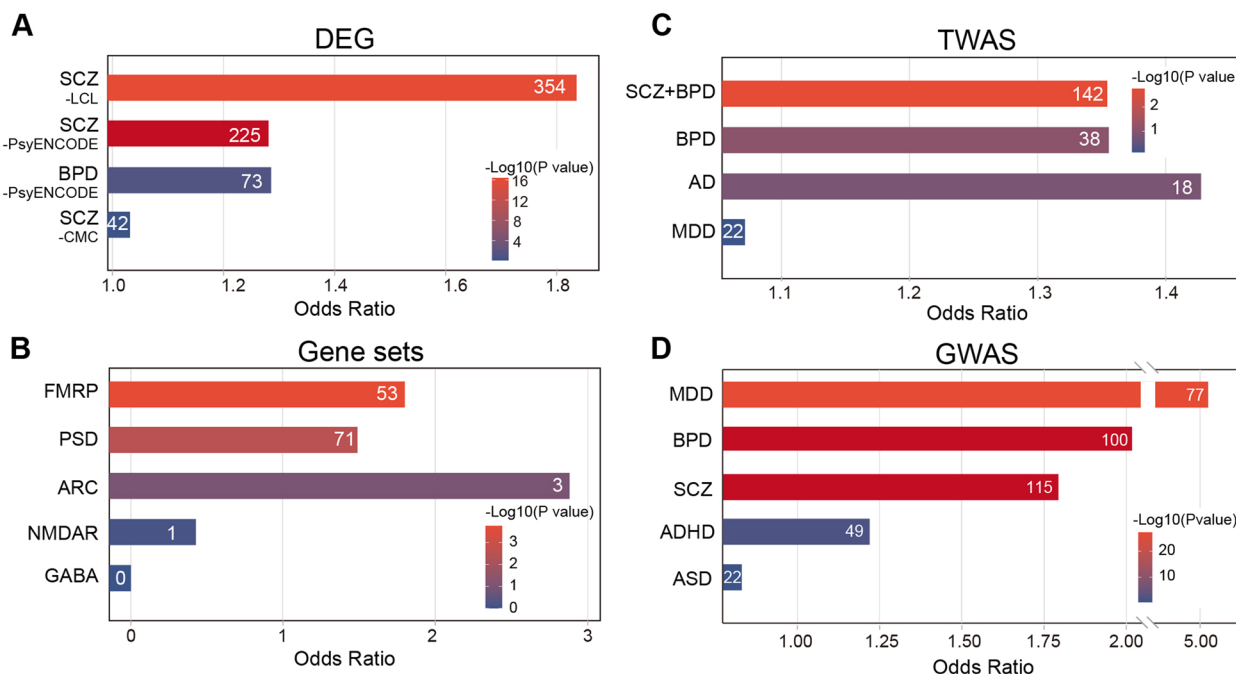


Fig 2 Disease-associated features of psyAShM genes. **A–D** Enrichment of 807 psyAShM sites in SCZ- or BPD-associated DEGs from the LCL, PsychENCODE, or CMC RNA-seq datasets (**A**), psychiatric disease-associated gene sets (**B**), TWAS (**C**), and GWAS (**D**) of SCZ, BPD, AD, MDD, ADHD, and ASD. The number of genes in each comparison is indicated

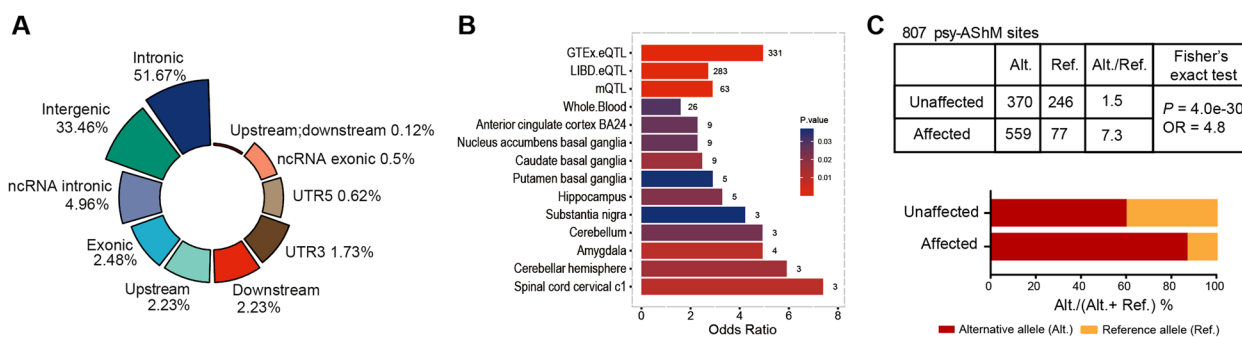


Fig 3 Characteristics of psyAShM sites. **A** The genomic context of 807 psyAShM sites. **B** Enrichment results of 807 psyAShM sites in eQTLs and brain mQTLs, and in genes expressed specifically in the brain or blood tissues. The number of genes in each comparison is indicated. **C** PsyAShM site counts showing preferential hydroxymethylation of alternative alleles in affected and unaffected individuals. The ratio is expressed as alternative/reference (Alt./Ref., upper) or alternative/(alternative+ reference) % (Alt./(Alt.+Ref.), lower)

significant association with psychiatric disorders. On average, alternative alleles of the 807 identified psyAShM sites were preferentially highly hydroxymethylated in affected (alt./ref.=7.3 and alt.%= 88%) compared to those in unaffected (alt./ref.=1.5 and alt.%= 60%) individuals (OR=4.8 and $P=4.0e-30$; Fig. 3C). These results suggested that psyAShM transition-induced hydroxymethylation alterations in alternative alleles might be associated with chromatin states, TF binding activity, and gene expression and ultimately affect the phenotype of SCZ/BPD.

Next, we examined the associations of psyAShM sites with chromatin state by examining the 807 psyAShM sites that overlapped with existing ChromHMM annotations in 10 brain tissues obtained from the NIH Roadmap epigenomics consortium [24, 25]. We observed that these psyAShM sites were enriched with most activated chromatin state-associated terms (genic enhancers, flanking active TSS, enhancers, strong transcription, and weak transcription) and only with two repressive chromatin state-associated terms (weak repression polyComb and quiescent/low; Fig. 4A). Moreover, we observed that DNase I hypersensitive sites (DHS), H3K27ac and H3K4me3 marks representing active chromatin states were more abundant on the allele with more DNA hydroxymethylation of the psyAShM sites in all regions, and significantly consistent patterns were shown for H3K27ac marks enriched in promoter regions (Fig. 4B). In contrast, repressive H3K9me3 signal was more abundant on the allele with less DNA hydroxymethylation of the psyAShM sites in all regions, including promoter regions. These results indicated that the psyAShM transitions might be involved in the activation of transcription.

We further evaluated the role of epiallele-specific TF binding at 807 psyAShM sites by focusing on a set of 358 TFs that had been assessed for binding affinity using

on the basis of the high-throughput systematic evolution of ligands determined by the exponential enrichment method [28]. We first identified 167 TFs that showed allele-specific binding affinity in at least three of 807 psyAShM sites and then examined the correlation between allelic differences in motif binding strength and the hydroxymethylation level of alternative alleles at these psyAShM loci for each of 167 TFs (Fig. 4C). We identified 5 TFs (PAX1, XBP1, SPDEF, NR4A2, and SOX9; orange in Fig. 4C) among the 11 TFs that showed individually significant correlations ($|r|>0.5$ and $P < 0.05$; Additional file 1: Table S7 and Additional file 2: Fig. S2) with larger differences in allele-specific TF binding affinities corresponding to higher hydroxymethylation levels of alternative alleles, and 6 TFs (POU3F2, NRF1, POU3F4, GLIS3, SP3, and MAX; red in Fig. 4C) showed negative correlations, with larger differences in allele-specific TF binding affinities corresponding to lower hydroxymethylation levels of alternative alleles. For example, POU3F2 showed alternative allele-specific binding affinity across four psyAShM sites, and a gain of hydroxymethylation at alternative alleles weakened its allele-specific binding affinities (Fig. 4D). Altogether, these results suggested the involvement of the psyAShM transition in gene expression alterations through its effects on epiallele-specific TF binding.

The rs4558409 regulatory locus displays epiallelic inhibition of POU3F2 binding affinity

We then sought to validate the downstream functional consequences of these psyAShM sites by predicting allelic differences in TF binding. We focused on 68 psyAShM sites embedded within the binding motifs of the 11 aforementioned TFs because each showed epiallele-specific binding affinities across these psyAShM sites (Fig. 4C). Of the 68 psyAShM sites, four sites found

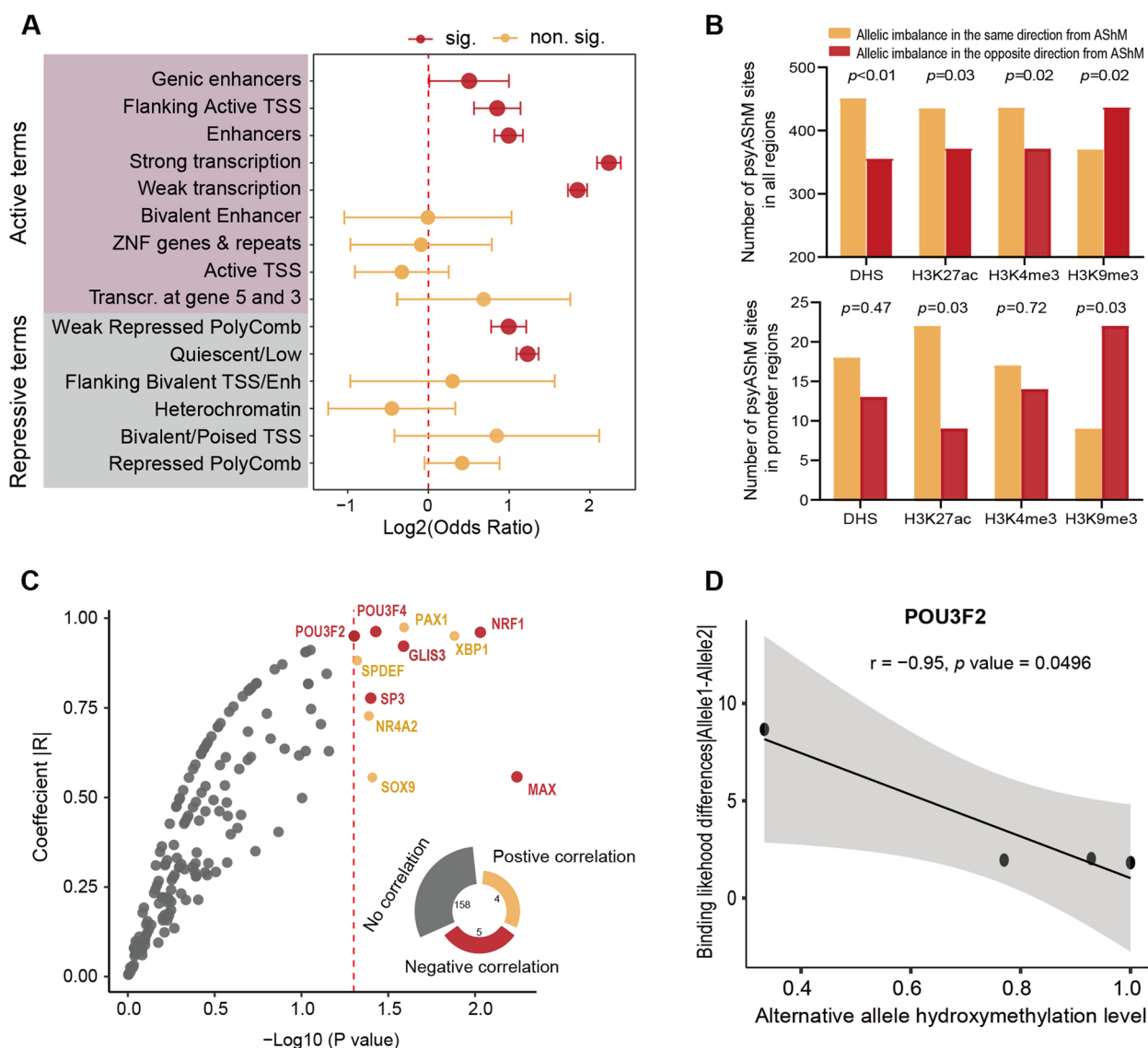


Fig 4 psyASHM sites exert epiallelic effects on chromatin states and TF binding affinities. **A** Enrichment of 807 psyASHM sites in chromatin states. $p < 0.05$ from Fisher's exact tests are shown with red points (\log_2 (OR) and 95% confidence intervals). **B** Number of psyASHM sites in the DNase I hypersensitive site (DHS) and histone marks overlapping ASHM loci throughout different classes of genomic elements. The P value from a binomial test is shown. **C** Effects of ASHM on absolute TF binding affinity based on position weight matrix scores (PWMs) at predicted TF binding sites. The negative correlation between absolute TF binding affinity differences and the alternative allele hydroxymethylation level at predicted TF binding sites is shown in red, the positive correlation between absolute TF binding affinity differences and the alternative allele hydroxymethylation level at predicted TF binding sites is shown in orange, and data showing no correlation are shown in gray. **D** Correlation between absolute POU3F2 binding affinity differences and alternative allele hydroxymethylation level at five predicted POU3F2 binding ASHM sites

to be eQTLs and subthreshold SCZ or BPD GWAS SNPs, exhibiting significant ASHM transitions in PDC twin pairs (Fig. 5A, B and Additional file 1: Table S8), were selected and assessed to validate their regulatory roles in SK-N-SH and HEK293T cell lines. We observed that three of the four SNPs, including *PLLP*-rs4558409 in the POU3F2 binding motif, *ITPKB*-rs708769 within the PAX1-binding motif, and *PLEC*-rs10866916 within

the SP3-binding motif, showed significant allelic effects on luciferase activities in both HEK293T and SK-N-SH cell lines (Fig. 5C). These allele-specific effects were consistent with eQTL patterns observed in the GTEx or CommonMind Consortium (CMC) datasets (Fig. 5B and Additional file 1: Table S8). Since *ITPKB*-rs708769 showed very low promoter activity and PAX1 also exhibited low expression in the GTEx dataset, we then

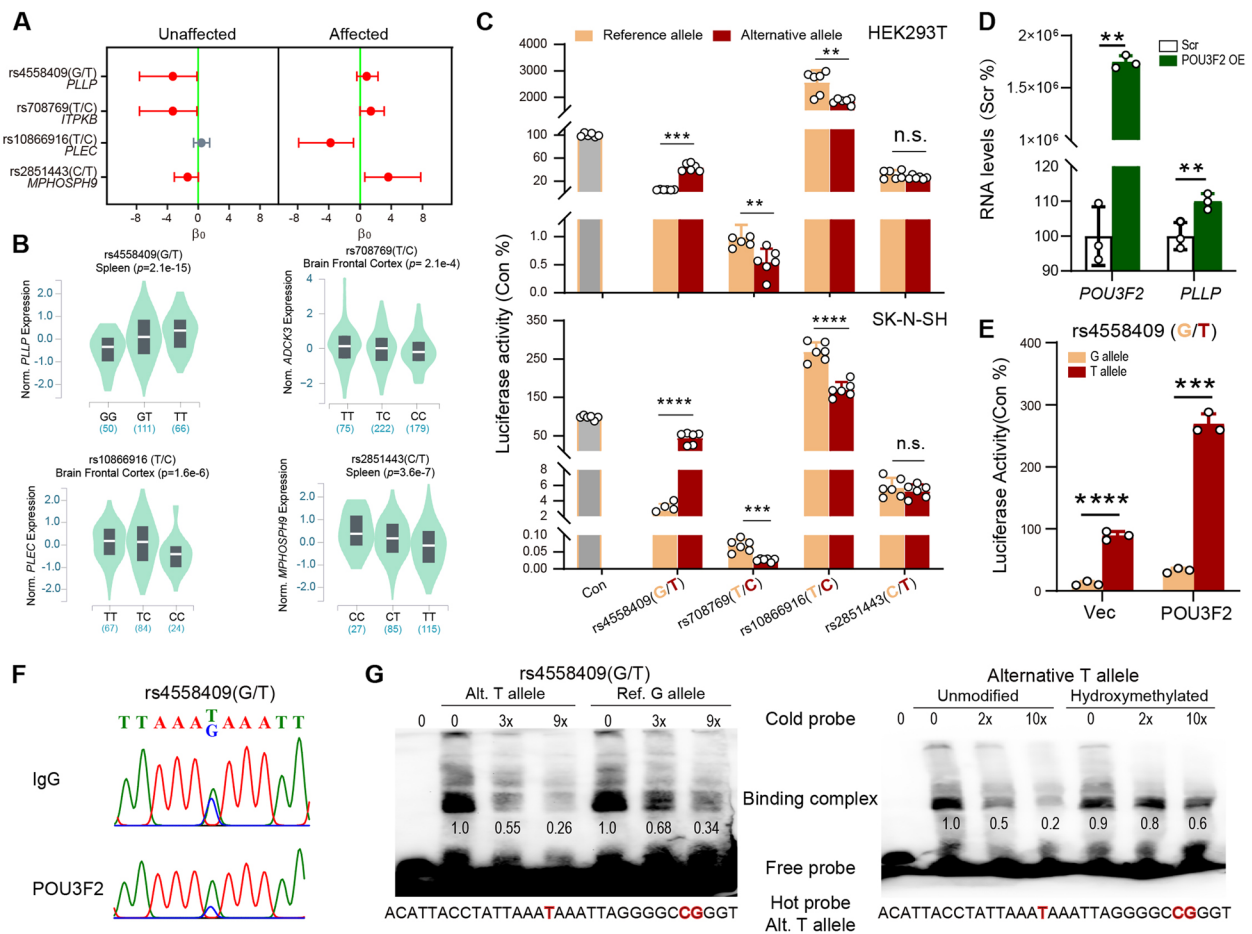


Fig 5 psyAshM transition exerts an allelic effect on TF binding affinity. **A** Estimates for four discordant ASHM sites in PDC twin pairs. Each point represents the β_0 value of the psyAshM sites showing the strongest evidence of ASHM in unaffected or affected individuals (highlighted in red, $p < 1e-3$) and a discordant ASHM pattern in PDC twin pairs ($BF > 10$). Error bars indicate 95% credible intervals. **B** eQTL patterns of the four psyAshM sites from the GTEx dataset. **C** Allelic effects of rs4558409 (G/T), rs708769 (T/C), rs10866916 (T/C), and rs2851443 (C/T) on promoter activity of the luciferase reporter in HEK293T and SK-N-SH cells are shown in columns with standard errors indicated by bars. The gray column represents cells transfected with pGL4.23-empty vector (Con), the orange column represents cells transfected with the reference allele-containing fragment cloned in pGL4.23, and the red column represents the alternative allele-containing fragment. **D, E** Effect of POU3F2 overexpression on endogenous *PLL*P expression (**D**) or rs4558409 allele-dependent promoter activity of the luciferase reporter (**E**) in HEK293T cells. Vec represents the pcDNA3.1 empty vector. **F** Sanger sequencing traces from ChIP-PCRs of POU3F2 and IgG at the rs4558409 (G/T) site in HEK293T cells with a heterozygous rs4558409 genotype. **G** EMSA and competition analysis. Assays were performed with the rs4558409 alternative T allele (red bold) as the hot probe, the rs4558409 alternative T allele or the reference G allele as the cold probe (left) and the alternative T allele without or with hydroxymethylated CpG (red bold) as another cold probe (right) with HEK293T nuclear extracts. Fold differences of molar excess of the cold probe compared to the hot probe and the relative intensity of the binding complex are shown underneath each panel. $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$ or ns, nonsignificant from *t* test

focused on the epiallele-specific effects of rs4558409 and rs10866916 on TF binding.

We observed that overexpression of *POU3F2* significantly increased the mRNA level of endogenous *PLL*P (Fig. 5D), and the alternative T allele rs4558409 exhibited greater activation of promoter activity than the reference G allele in luciferase reporters cotransfected with *POU3F2* into HEK293T cells (Fig. 5E). This result is consistent with rs4558409 showing an alternative T

allele-specific *POU3F2* binding affinity (Additional file 1: Table S8) and a T-allele-dependent upregulation of *PLL*P RNA, as determined based on eQTL patterns (Fig. 5B). Although we observed that SP3 showed upregulation of endogenous *PLEC* transcription and rs10866916 reference T-allele-specific promoter activity enhancement in a luciferase reporter assay after cotransfection with *SP3* in HEK293T cells (Additional file 2: Fig. S3), this finding was inconsistent with our TF binding affinity prediction

results, which revealed alternative C-allele-specific SP3-binding affinity (Additional file 1: Table S8), indicating that other inhibitory TFs were involved in regulating *PLEC* expression through the rs10866916 site. We further verified that POU3F2 displayed binding activity around the rs4558409 (G/T) site (Additional file 2: Fig. S4) and showed a preference for occupying the alternative T allele over the reference G allele in POU3F2-ChIP-DNA compared with IgG-ChIP-DNA in the HEK293T cell line with heterozygous genotypes at rs4558409 (Fig. 5F). Moreover, we observed that the rs4558409 alternative T allele displayed higher binding activity than the reference G allele (Fig. 5G left), and a hydroxymethylated T allele probe displayed a lower binding activity than an unmodified probe (Fig. 5G right) and a methylated probe (Additional file 2: Fig. S5), as demonstrated via EMSA with HEK293T nuclear extracts. The AShM transition toward hyper-hydroxymethylation correlated with the rs4558409 alternative T allele under disease conditions might contribute to reduced *PLLP* expression through the inhibition of POU3F2 binding affinity. Intriguingly, rs4558409 is located in the intronic regions of *PLLP*, which has been reported to be significantly downregulated in post-mortem SCZ brains ($\text{Log}_2\text{FC} = -0.129$, $p = 0.00058$), and marginally in BPD patients ($\text{Log}_2\text{FC} = -0.104$, $p = 0.051$), compared with that in nonpsychiatric controls in the PsychENCODE brain RNA-seq datasets [47], supporting the putative inhibitory effects of AShM transitions at

rs4558409 on dysregulation of *PLLP* expression in SCZ or BPD.

Disruption of rs4558409 promotes neural development and vesicle trafficking

To examine the function of the rs4558409-containing region, we employed Cas9/sgRNA editing to disrupt the sequences around rs4558409 in human neuroblastoma SK-N-SH cells (rs4558409-KO; Additional file 2: Fig. S6), which has the ability to differentiate into a neuronal phenotype characterized by extensive neurite outgrowth and is the most widely used system to study neurodevelopment [3, 41–45]. We observed that rs4558409-KO significantly increased *PLLP* mRNA expression levels (increased by 475%; Fig. 6A) and protein expression levels (by 169%; Fig. 6B) in SK-SN-SH cells. *PLLP* encodes the proteolipid plasmalogen, which is a main component of synaptic plasma membranes and myelin sheaths and is involved in intracellular transport and neurite growth [49]. We then observed that rs4558409-KO promoted all-trans retinoic acid (ATRA)-induced neural development (Fig. 6C), causing an increase in the neurite-like length of SK-N-SH cells after treatment with 10 μM ATRA (Fig. 6D). Next, we monitored exocytosis using the dye FM4-64 in the presence of 75 mM KCl to examine whether rs4558409-KO has a direct effect on vesicle secretion from KCl-induced SK-N-SH cells (Additional file 2: Fig. S7A). We observed that FM4-64 was depleted

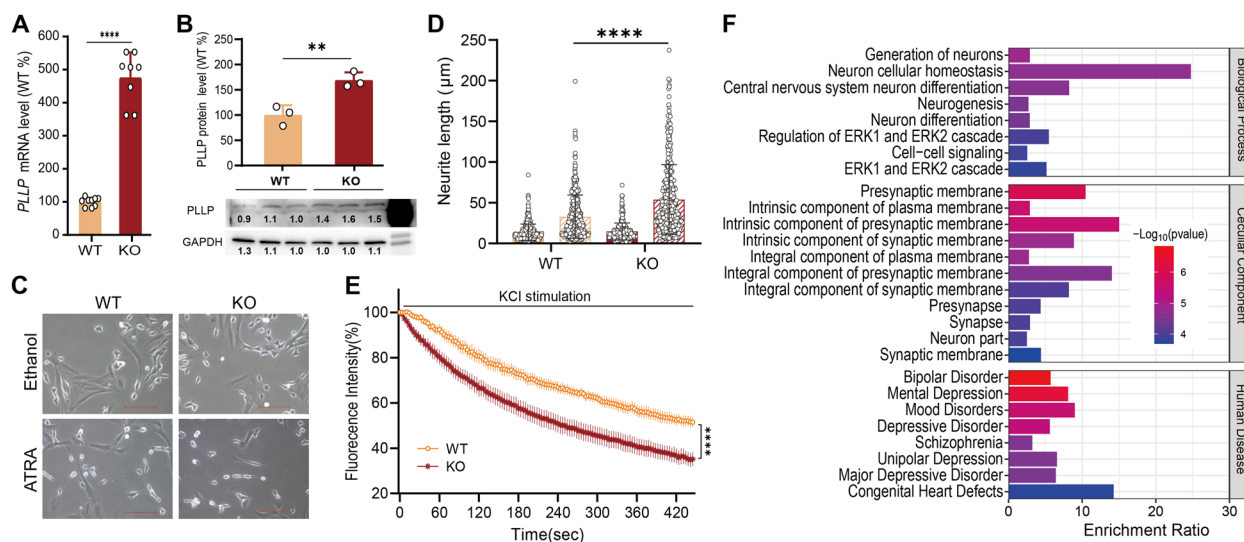


Fig 6 Disruption of rs4558409 promotes neural development and vesicle trafficking. **A, B** Effects of rs4558409-KO (KO) on *PLLP* RNA (**A**) and protein (**B**) expression levels in SK-N-SH cells. $**p < 0.01$ and $****p < 0.0001$ from two-tailed Student's *t* test. The relative intensity of binding as determined via Western blot analysis is shown underneath each band. **C** rs4558409-KO promoted neural development of SK-N-SH cells without or with ATRA treatment. The scale bar represents 100 μm . **D** Neurite-like length of the wild-type (WT, orange) or rs4558409-KO (KO, red) SK-N-SH cells without (solid) or with ATRA treatment (hatched). $****p < 0.0001$ from one-way ANOVA with Tukey's multiple comparisons test. **E** FM4-64 imaging analysis of the wild-type (WT, orange) or rs4558409-KO (KO, red) SK-N-SH cells. $****p < 0.0001$ obtained from two-way ANOVA. **F** Functional enrichment analysis of rs4854158-KO-induced DEGs in SK-N-SH cells

faster in rs4558409-KO cells than in unperturbed SK-N-SH cells (Fig. 6E and Additional file 2: Fig. S7B), indicating enhanced exocytosis driven by rs4558409-KO.

We also examined the transcriptome profiles of rs4854158-KO SK-N-SH cells without ATRA treatment by RNA-seq and identified 1115 DEGs ($|\text{Log}_2\text{FC}| > 1$ and $\text{FDR} < 0.1$) including 56 upregulated DEGs and 59 downregulated DEGs in rs4854158-KO cells (Additional file 1: Table S9). Functional enrichment analysis revealed that GO-BP such as generation of neurons, central nervous system neuron differentiation, neurogenesis, and neuron differentiation; GO-Cellular Component such as presynaptic membrane, synapse, neuron part, and synaptic membrane; and human diseases such as BPD, SCZ, and MDD were significantly enriched among rs4854158-KO-induced DEGs (Fig. 6F and Additional file 1: Table S10). Although *PLLP* did not show a significant increase ($\text{Log}_2\text{FC} = 0.2289$, $p = 0.6264$) in rs4854158-KO cells, the upregulated *ERBB4* (Erb-B2 Receptor Tyrosine Kinase 4) and *NRG1* (Neuregulin 1) and the downregulated *GADI* (Glutamate Decarboxylase 1) in rs4854158-KO cells, which are involved in neural development, may partially explain the altered neural development in rs4854158-KO cells. These genes have also been reported to associated with SCZ or BPD [50, 51].

Discussion

Here, we performed a genome-wide examination of ASHM sites from MZ twins and identified a large number of SNPs showing ASHM transitions across MZ twins with discordant phenotypes. These psyASHM sites displayed epiallele-specific effects on chromatin states and TF binding, and their ASHM transitions may have led to dysregulated gene expression and functions and may have eventually increased the risk of SCZ or BPD. We then employed multiple lines of data to show that competitive binding of POU3F2 on the alternative T allele at the psyASHM site rs4558409 (G/T) can enhance *PLLP* expression, while the hydroxymethylated alternative allele alleviating the POU3F2 binding activity at the rs4558409 site might be associated with the downregulated *PLLP* expression observed in patients with BPD/SCZ. This study has etiological implications for the ASHM transition in patients with SCZ and BPD.

Our findings showed that sensitivity of the hydroxymethylome to genetic variation among MZ twin pairs with discordant phenotypes displayed disease-associated features through epiallele-specific effects on chromatin states, TF binding, and gene expression. Our allelic epigenome profiling of 17 MZ twin pairs revealed that 1.2% of informative SNPs showed ASHM ($\text{FDR} < 0.1$), which was a more conservative outcome than estimates for ASM (2.2%) in the same twin pairs analyses in our previous study [3]. Some of our identified ASHM sites

(11.7%) displayed sequence-dependent ASM, as shown in our previous study [3], and were located in previously reported imprinted regions. The regulatory role of ASHM sites was shown by the evidence of their uneven distribution across the genome with 90% of the psyASHM sites located in noncoding regions, greater enrichment of psyASHM sites in the active chromatin state, and ASHM transitions that may induce changes in chromatin state and TF binding ability, eventually inducing altered gene expression. The regulatory roles of psyASHM were further demonstrated by their significant enrichment in brain eQTLs/mQTLs, SCZ/BPD-associated DEGs, and TWAS or GWAS of SCZ/BPD. In fact, we recently reported that ASM sites are more enriched in repressed chromatin states and showed epiallelic effects on TF binding and gene expression [3]. DNA methylation has long been associated with gene silencing, inferring that DNA demethylation can lead to gene activation. As an intermediate of DNA demethylation processing, 5hmC has also been reported to be highly enriched in the gene bodies of transcriptionally active genes, promoters, and enhancers and to undergo highly dynamic changes during development and differentiation and in the context of neuropsychiatric disorders [10, 12]. The functional implication of sequence-dependent ASM and ASHM sites examined in our study, together with mostly functional genetic variants identified in noncoding regions via GWAS [2], provided further evidence for dynamic DNA methylation/demethylation in mediating associations among genetic risk burden, environmental or epigenetic risk exposure, and phenotype.

Our study validated the biological role of ASHM transitions at the *PLLP* rs4558409 locus, leading to alterations in POU3F2 binding ability and contributing to the downregulation of *PLLP*, providing etiological implications for SCZ or BPD. *PLLP* encodes the proteolipid plasmalogen, which is a main component of synaptic plasma membranes and myelin sheaths and is involved in intracellular transport and neurite growth [49]. ASHM transitions at the *PLLP* rs4558409 locus across PDC twin pairs reduced *PLLP* transcriptional activity through hydroxymethylation-mediated inhibition of active POU3F2 binding affinity at the active T allele of the rs4558409 (G/T) regulatory locus. The allele-specific regulatory role of rs4558409 in a luciferase reporter assay was consistent with the eQTL pattern. The alternative T allele of rs4558409 showing higher binding activity was demonstrated via ChIP and EMSAs, and its hyper-hydroxymethylation/methylation, which reduced binding ability, might significantly contribute to the reduced *PLLP* expression observed in postmortem SCZ brain [47], and rs4558409 disruption promotes neural development and vesicle trafficking. Similarly, reduced *PLLP* has been observed in the temporal cortex of patients with MDD [52]. In fact,

rs4558409 is a subthreshold GWAS SNP associated with SCZ ($p=0.007479$ in PGC3) [36] and MDD ($p=0.026$) [40], and hyper-hydroxymethylation at the risk alternative T allele of rs4558409 (G/T), which reduced TF binding affinity and *PLLP* expression, might increase disease susceptibility. *PLLP* plays an important role in membrane biogenesis and myelination [49], and distorted oligodendrocyte differentiation and subsequent defects in myelination might lead to SCZ and BPD onset [49, 53]. Intriguingly, *POU3F2* is one of the key regulators coexpressed with many genes within the SCZ-related module [54], and *POU3F2*-regulated target genes may contribute to neurodevelopment and synaptic function in various ways [55]. In addition, *POU3F2*-deficient mice exhibited impaired hippocampal neurogenesis [56]. *POU3F2* expression is decreased in cerebral organoids in patients with BPD [57], and its expression is associated with differential SCZ-associated alterations in brain tissues, including downregulation in the SZDB dataset [58] but upregulation in the PsychENCODE dataset [47]. In summary, involvement of ASHM transitions at rs4558409 regulatory loci might contribute to disease-associated *PLLP* downregulation through epiallele-specific inhibition of *POU3F2* binding.

This study also has several limitations. First, although our study identified that disruption of rs4558409-contained region promotes neural development and vesicle trafficking, whether rs4558409 (G/T) display allelic effects on neural development and vesicle trafficking remains to be further examined through single-base mutation strategy in SK-N-SH cells. Second, whether KO of *PLLP* could restore neural development induced by rs4558409-KO may be needed to validate the role of rs4558409 in regulation of *PLLP*. Third, whether *ERBB4*, *NRG1*, or *GAD1* except for *PLLP* are regulated by rs4854158 may provide further mechanism of rs4854158 on regulation of neural development since expression of those genes are significantly dysregulated in rs4558409-KO cells. Finally, despite the human neuroblastoma SK-N-SH cell lines employed in this study are the most widely applied and cited in vitro system for neurodevelopmental and neuropsychiatric studies, further studies in other appropriate test models to validate the function of rs4854158 in pathogenesis of the BPD or SCZ are needed.

Conclusions

We observed that the allelic imbalances in hydroxymethylation vary across MZ twin pairs with discordant phenotypes were associated with SCZ/BPD-associated features by altering gene regulation, contributing to our understanding of the interaction between genetic and epigenetic factors in mediating individual differences in disease susceptibility and providing a powerful strategy for prioritizing SCZ/BPD-associated genes or variants.

Abbreviations

5hmC	5-Hydroxymethylcytosine
AD	Alzheimer disease
AShM	Allele-specific hydroxymethylation
ASM	Allele-specific methylation
ASoC	Allele-specific open chromatin
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
ATRA	All-trans retinoic acid
BCC	BPD-concordant twins
BDC	BPD-discordant twins
BF	Bayes factors
BPD	Bipolar disorder
ChIP	Chromatin immunoprecipitation
CMC	CommonMind Consortium
DEG	Differentially expressed genes
EMSA	Electrophoretic mobility shift assay
eQTL	Expression quantitative trait loci
ESCs	Embryonic stem cells
FC	Fold-changes
FMRP	Fragile X mental retardation protein
GTE _x	Genotype-Tissue Expression
GWAS	Genome-wide association study
HCC	Healthy concordant twins
KO	Knock out
LCLs	Lymphoblastoid cell lines
LIBD	Lieber institute for brain development
MDD	Major depressive disorder
MZ	Monozygotic
OR	Odds ratio
PCC	Psychiatric disorder-concordant twins
PDC	Psychiatric disorders in discordant MZ twins
<i>PLLP</i>	Plasmolipin
<i>POU3F2</i>	POU Class 3 Homeobox 2
PSD	Postsynaptic density
psyASHM sites	Psychiatric disorder-associated ASHM sites
SCC	SCZ-concordant twins
SCZ	Schizophrenia
SDC	SCZ-discordant twins
SNPs	Single-nucleotide polymorphisms
TET	Ten eleven translocation
TFs	Transcription factors
TSS	Transcription start sites
TWAS	Transcriptome-wide association study
WGS	Whole-genome sequencing

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-023-03177-y>.

Additional file 1: Table S1. Summary of 5hmC-seq and WGS data for each twin pair. **Table S2.** Primers and probes employed in this study. **Table S3.** The number of ASHM sites identified in MZ twins. **Table S4.** Detailed information about 534 ASHM sites in 53 previous reported known imprinted clusters. **Table S5.** Information about 807 ASHM sites. **Table S6.** Functional enrichment of GO biological process (BP) annotations among 807 ASHM sites annotated genes. **Table S7.** Information of allele-specific transcript factor (TF) binding affinity of 807 psyASHM sites. **Table S8.** Detailed information of four ASHM sites. **Table S9.** rs4558409-KO induced DEGs in SK-N-SH cells. **Table S10.** GO-BP, GO-CC and human disease enrichment results of rs4558409-KO induced DEGs.

Additional file 2: Fig. S1. Genome-wide DNA hydroxymethylation patterns of MZ twins. **Fig. S2.** Correlation analysis results. **Fig. S3.** Allelic effects of rs10866916 on the SP3 activity. **Fig. S4.** ChIP-qPCR. **Fig. S5.** EMSA and competition analysis. **Fig. S6.** Sequencing chromatographs of PCR products of rs4558409-dirupted *PLLP* in SK-N-SH cells. **Fig. S7.** FM4-64 imaging analysis of neurite length.

Acknowledgements

We thank the individuals for their willingness to participate in this study. We thank the Central Laboratory of Southern Medical University for providing platform.

Authors' contributions

Conceived and designed the experiments: JY, ZH, PJ, JHC, and CZ. Performed the experiments: JY, ZH, QL, YtL, CYR, and SL. Analyzed the data: JY, ZH, ZL, ZW, XW, YL, JL, PJ, JHC, and CZ. Collected and diagnosed the subjects: TJ, QY, and MJ. Wrote the paper: JY, ZH, and CZ. JY and ZH contributed equally to this study. All authors read and approved the final manuscript.

Funding

This work was supported by the funds from the National Natural Science Foundation of China [grant numbers 82171502, 82201655, 82001411, 82002237, 82101577, 32000419, and 32000419]; China Postdoctoral Science Foundation [grant number 2022M721506 and 2020M682806] and the Guangdong Science and Technology Foundation [grant numbers 2022A1515011042 and 2023A1515012374].

Availability of data and materials

All the data generated in this study were shown in the main text and additional files. RNA sequencing data of rs4854158-KO SK-NSH cells have been deposited to the NCBI GEO site with the GEO number GSE246593. Additional information is also available upon reasonable request to the corresponding authors.

Declarations

Ethics approval and consent to participate

All participants in this study provided informed consent prior to the study following presentation of the nature of the procedures. Approval for the extended cohort study was obtained as outlined by the protocol approved by Medical Ethics Committee of Zhujiang Hospital of Southern Medical University (#2022-KY-086) and the third People's Hospital of Zhongshan (SSYLL20210301). The study was conducted in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Key Laboratory of Mental Health of the Ministry of Education, Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence, Guangdong Province Key Laboratory of Psychiatric Disorders, and Department of Medical Genetics, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China. ²Department of Rehabilitation, Zhujiang Hospital of Southern Medical University, Guangzhou, China. ³Department of Psychiatry, the Affiliated Brain Hospital of Guangzhou Medical University (Guangzhou Huiai Hospital), Guangzhou, Guangdong, China. ⁴The Third People's Hospital of Zhongshan, Zhongshan, Guangdong, China. ⁵Departments of Human Genetics, Emory University, Atlanta, GA, USA. ⁶Laboratory of Genomic and Precision Medicine, Wuxi School of Medicine, Jiangnan University, Wuxi, China. ⁷Guangdong Provincial People's Hospital (Guangdong Academy of Medical Science), Guangdong Mental Health Center, Southern Medical University, Guangzhou, China. ⁸Experimental Education/Administration Center, School of Basic Medical Science, Southern Medical University, Guangzhou, China.

Received: 5 September 2023 Accepted: 13 November 2023

Published online: 12 December 2023

References

- Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*. 2009;373(9659):234–9.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190–5.
- Li Q, Wang Z, Zong L, Ye L, Ye J, Ou H, Jiang T, Guo B, Yang Q, Liang W, et al. Allele-specific DNA methylation maps in monozygotic twins discordant for psychiatric disorders reveal that disease-associated switching at the EIPR1 regulatory loci modulates neural function. *Mol Psychiatry*. 2021;26(11):6630–42.
- Onuchic V, Lurie E, Carrero I, Pawliczek P, Patel RY, Rozowsky J, Galeev T, Huang Z, Altshuler RC. Zhang Z et al: Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci. *Science*. 2018;361(6409):eaar3146.
- Meaburn EL, Schalkwyk LC, Mill J. Allele-specific methylation in the human genome: implications for genetic studies of complex disease. *Epigenetics*. 2010;5(7):578–82.
- Kadota M, Yang HH, Hu N, Wang C, Hu Y, Taylor PR, Buetow KH, Lee MP. Allele-specific chromatin immunoprecipitation studies show genetic influence on chromatin state in human genome. *PLoS Genet*. 2007;3(5):e81.
- Zhang S, Zhang H, Zhou Y, Qiao M, Zhao S, Kozlova A, Shi J, Sanders AR, Wang G, Luo K, et al. Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science*. 2020;369(6503):561–5.
- Kriaucionis S, Heintz N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science*. 2009;324(5929):929–30.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*. 2009;324(5929):930–5.
- Jiang T, Zong L, Zhou L, Hou Y, Zhang L, Zheng X, Han H, Li S, Zhang W, Zhang J, et al. Variation in global DNA hydroxymethylation with age associated with schizophrenia. *Psychiatry Res*. 2017;257:497–500.
- Zong L, Zhou L, Hou Y, Zhang L, Jiang W, Zhang W, Wang L, Luo X, Wang S, Deng C, et al. Genetic and epigenetic regulation on the transcription of *GABRB2*: Genotype-dependent hydroxymethylation and methylation alterations in schizophrenia. *J Psychiatr Res*. 2017;88:9–17.
- Kuehner JN, Chen J, Bruggeman EC, Wang F, Li Y, Xu C, McEachin ZT, Li Z, Chen L, Hales CM, et al. 5-hydroxymethylcytosine is dynamically regulated during forebrain organoid development and aberrantly altered in Alzheimer's disease. *Cell reports*. 2021;35(4): 109042.
- He B, Zhang C, Zhang X, Fan Y, Zeng H, Liu J, Meng H, Bai D, Peng J, Zhang Q, et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat Commun*. 2021;12(1):4249.
- Ni C, Jiang W, Wang Z, Wang Z, Zhang J, Zheng X, Liu Z, Ou H, Jiang T, Liang W et al: LncRNA-AC006129.1 reactivates a SOCS3-mediated anti-inflammatory response through DNA methylation-mediated CIC downregulation in schizophrenia. *Mol Psychiatry*. 2021;26(8):4511–28.
- Liang W, Hou Y, Huang W, Wang Y, Jiang T, Huang X, Wang Z, Wu F, Zheng J, Zhang J et al: Loss of schizophrenia-related miR-501-3p in mice impairs sociability and memory by enhancing mGluR5-mediated glutamatergic transmission. *Sci Adv*. 2022;8(33):eabn7357.
- Bell JT, Spector TD. A twin approach to unraveling epigenetics. *Trends in Genetics*. 2011;27(3):116–25.
- Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen C-H, Zhang W, Jian X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*. 2011;29(1):68–72.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012;9(4):357–U354.
- Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics*. 2014;30(2):284–6.
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568–76.
- Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009;71(2):319–92.

22. McCoy RC, Wakefield J, Akey JM. Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. *Cell*. 2017;168(5):916–927 e912.
23. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research*. 2009;37:W305–11.
24. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317–30.
25. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28(10):1045–8.
26. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nature genetics*. 2018;50(8):1171–9.
27. Coetzee SG, Coetzee GA, Hazelett DJ. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics*. 2015;31(23):3847–9.
28. Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, et al. DNA-binding specificities of human transcription factors. *Cell*. 2013;152(1–2):327–39.
29. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, Moran J, Chambert K, Toncheva D, Georgieva L, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry*. 2012;17(2):142–53.
30. Darnell JC, Van Driesche SJ, Zhang C, Hung KY, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW, et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*. 2011;146(2):247–61.
31. Bayes A, van de Lagemaat LN, Collins MO, Croning MD, Whittle IR, Choudhary JS, Grant SG. Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci*. 2011;14(1):19–21.
32. Pocklington AJ, Rees E, Walters JT, Han J, Kavanagh DH, Chambert KD, Holmans P, Moran JL, McCarrroll SA, Kirov G, et al. Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia. *Neuron*. 2015;86(5):1203–14.
33. Gandalf MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, Won H, van Bakel H, Varghese M, Wang Y et al: Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362(6420):eaat8127.
34. Hoffman GE, Bendl J, Voloudakis G, Montgomery KS, Sloofman L, Wang YC, Shah HR, Hauberg ME, Johnson JS, Girdhar K, et al. CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci Data*. 2019;6(1):180.
35. Duan J, Goring HHH, Sanders AR, Moy W, Freda J, Drigalenko EI, Kos M, He D, Gejman PV. Mgs: Transcriptomic signatures of schizophrenia revealed by dopamine perturbation in an ex vivo model. *Translational psychiatry*. 2018;8(1):158.
36. Trubetskoy V, Pardini AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, Bryois J, Chen CY, Dennison CA, Hall LS, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502–8. <https://doi.org/10.1038/s41586-022-04434-5>.
37. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, Als TD, Bigdeli TB, Borte S, Bryois J, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet*. 2021;53(6):817–29. <https://doi.org/10.1038/s41588-021-00857-4>.
38. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, Baldursson G, Belliveau R, Bybjerg-Grauholm J, Baekvad-Hansen M, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. 2019;51(1):63–75.
39. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, Pallesen J, Agerbo E, Andreassen OA, Anney R, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431–44.
40. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, Coleman JRI, Hagenaars SP, Ward J, Wigmore EM, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343–52.
41. Hecht PM, Ballesteros-Yanez I, Grepo N, Knowles JA, Campbell DB. Noncoding RNA in the transcriptional landscape of human neural progenitor cell differentiation. *Front Neurosci*. 2015;9:392.
42. Kaushik G, Xia Y, Yang L, Thomas MA. Psychoactive pharmaceuticals at environmental concentrations induce in vitro gene expression associated with neurological disorders. *BMC Genomics*. 2016;17 Suppl 3(Suppl 3):435.
43. Kovalevich J, Langford D. Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology. *Methods Mol Biol*. 2013;1078:9–21.
44. Kaushik G, Xia Y, Pfau JC, Thomas MA. Dysregulation of autism-associated synaptic proteins by psychoactive pharmaceuticals at environmental concentrations. *Neurosci Lett*. 2017;661:143–8.
45. Pizzi M, Boroni F, Bianchetti A, Moraitis C, Sarnico I, Benarese M, Goffi F, Valerio A, Spano P. Expression of functional NR1/NR2B-type NMDA receptors in neuronally differentiated SK-N-SH human cell line. *Eur J Neurosci*. 2002;16(12):2342–50.
46. Siegert S, Seo J, Kwon EJ, Rudenko A, Cho S, Wang W, Flood Z, Martorell AJ, Ericsson M, Mungenast AE, et al. The schizophrenia risk gene product miR-137 alters presynaptic plasticity. *Nat Neurosci*. 2015;18(7):1008–16.
47. Gandalf MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, Won H, van Bakel H, Varghese M, Wang Y, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362(6420):1265.
48. Bipolar D, Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell*. 2018;173(7):1705–1715 e1716 Electronic address drive, Bipolar D, Schizophrenia Working Group of the Psychiatric Genomics C.
49. Shulgin AA, Lebedev TD, Prassolov VS, Spirin PV. Plasmolipin and Its Role in Cell Processes. *Mol Biol*. 2021;55(6):773–85.
50. Hahn CG, Wang HY, Cho DS, Talbot K, Gur RE, Berrettini WH, Bakshi K, Kamins J, Borgmann-Winter KE, Siegel SJ, et al. Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. *Nature medicine*. 2006;12(7):824–8.
51. Ruzicka WB, Subburaju S, Benes FM. Circuit- and Diagnosis-Specific DNA Methylation Changes at γ -Aminobutyric Acid-Related Genes in Postmortem Human Hippocampus in Schizophrenia and Bipolar Disorder. *JAMA Psychiatry*. 2015;72(6):541–51.
52. Aston C, Jiang L, Sokolov BP. Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients with major depressive disorder. *Mol Psychiatry*. 2005;10(3):309–22.
53. Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, Jones PB, Starkey M, Webster MJ, Yolken RH, Bahn S. Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet*. 2003;362(9386):798–805.
54. Chen C, Meng Q, Xia Y, Ding C, Wang L, Dai R, Cheng L, Gunaratne P, Gibbs RA, Min S et al: The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. *Sci Transl Med*. 2018;10(472):eaat8178.
55. Ding C, Zhang C, Kopp R, Kuney L, Meng Q, Wang L, Xia Y, Jiang Y, Dai R, Min S, et al. Transcription factor POU3F2 regulates TRIM8 expression contributing to cellular functions implicated in schizophrenia. *Molecular Psychiatry*. 2021;26(7):3444–60.
56. Hashizume K, Yamanaka M, Ueda S. POU3F2 participates in cognitive function and adult hippocampal neurogenesis via mammalian-characteristic amino acid repeats. *Genes, brain, and behavior*. 2018;17(2):118–25.
57. Kathuria A, Lopez-Lengowski K, Vater M, McPhie D, Cohen BM, Karmacharya R. Transcriptome analysis and functional characterization of cerebral organoids in bipolar disorder. *Genome medicine*. 2020;12(1):34.
58. Wu Y, Yao YG, Luo XJ. SZDB: A Database for Schizophrenia Genetic Research. *Schizophr Bull*. 2017;43(2):459–71.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.