

RESEARCH

Open Access



# Functional genomic analysis delineates regulatory mechanisms of GWAS-identified bipolar disorder risk variants

Rui Chen<sup>1,2†</sup>, Zhihui Yang<sup>1,2†</sup>, Jiewei Liu<sup>1†</sup>, Xin Cai<sup>1,2†</sup>, Yongxia Huo<sup>1</sup>, Zhijun Zhang<sup>3,4</sup>, Ming Li<sup>1\*</sup>, Hong Chang<sup>1\*</sup> and Xiong-Jian Luo<sup>1,3,5\*</sup>

## Abstract

**Background:** Genome-wide association studies (GWASs) have identified multiple risk loci for bipolar disorder (BD). However, pinpointing functional (or causal) variants in the reported risk loci and elucidating their regulatory mechanisms remain challenging.

**Methods:** We first integrated chromatin immunoprecipitation sequencing (ChIP-Seq) data from human brain tissues (or neuronal cell lines) and position weight matrix (PWM) data to identify functional single-nucleotide polymorphisms (SNPs). Then, we verified the regulatory effects of these transcription factor (TF) binding–disrupting SNPs (hereafter referred to as “functional SNPs”) through a series of experiments, including reporter gene assays, allele-specific expression (ASE) analysis, TF knockdown, CRISPR/Cas9-mediated genome editing, and expression quantitative trait loci (eQTL) analysis. Finally, we overexpressed *PACS1* (whose expression was most significantly associated with the identified functional SNPs rs10896081 and rs3862386) in mouse primary cortical neurons to investigate if *PACS1* affects dendritic spine density.

**Results:** We identified 16 functional SNPs (in 9 risk loci); these functional SNPs disrupted the binding of 7 TFs, for example, CTCF and REST binding was frequently disrupted. We then identified the potential target genes whose expression in the human brain was regulated by these functional SNPs through eQTL analysis. Of note, we showed dysregulation of some target genes of the identified TF binding–disrupting SNPs in BD patients compared with controls, and overexpression of *PACS1* reduced the density of dendritic spines, revealing the possible biological mechanisms of these functional SNPs in BD.

<sup>†</sup>Rui Chen, Zhihui Yang, Jiewei Liu and Xin Cai are equally contributed to this work.

\*Correspondence: limingkiz@mail.kiz.ac.cn; changhong@mail.kiz.ac.cn; luoxiongjian@mail.kiz.ac.cn

<sup>1</sup> Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China  
<sup>5</sup> Zhongda Hospital, School of Life Sciences and Technology, Advanced Institute for Life and Health, Southeast University, Nanjing, Jiangsu 210096, China

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



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusions:** Our study identifies functional SNPs in some reported risk loci and sheds light on the regulatory mechanisms of BD risk variants. Further functional characterization and mechanistic studies of these functional SNPs and candidate genes will help to elucidate BD pathogenesis and develop new therapeutic approaches and drugs.

**Keywords:** Bipolar disorder, Genome-wide association study (GWAS), Functional genomics, Transcription factor-disrupting SNPs, Expression quantitative trait loci (eQTL), Regulatory mechanisms

## Background

Bipolar disorder (BD) is a severe mental disorder that affects the emotion, cognition, and behaviour of affected individuals, and it affects more than 1% of the world's population [1, 2]. BD is characterized by recurrent alteration between hypomania and depression, and this disorder is classified into two major clinical subtypes: bipolar disorder type I (BD-I) (which is characterized by hypermanic symptoms) and bipolar disorder type II (BD-II) (which is mainly characterized by hypomanic episodes and severe depression episodes [1]). BD is associated with a high risk of morbidity [3, 4] and mortality (the suicide rate of individuals with BD is approximately 20–30 times higher than that of the general population [5]), which makes it a leading cause of disability worldwide.

Although the pathogenesis of BD remains to be elucidated, converging evidence suggests that both genetic and environmental factors are involved [6, 7]. Environmental risk factors, such as a lack of social support [8], life stress, and sleep-wake cycle disruption, have been reported to have a role in BD [9]. In addition, the high heritability (approximately 80%) indicates the major role of genetic components in BD [10–12]. Over the past decade, several BD risk loci have been identified by genome-wide association studies (GWASs) [13–16]. Despite the great success of these GWASs, to date, the genetic mechanisms of BD (i.e. how risk variants confer risk for BD) remain largely unknown. Considering that most BD risk variants are located in noncoding regions, it is likely that these risk variants confer risk for BD by regulating gene expression [17]. However, pinpointing the functional variants (in the reported risk loci) and elucidating their roles in BD remain major challenges (due to the complexity of linkage disequilibrium (LD) and gene regulation).

To highlight the functional (or causal) risk variants (in the reported BD risk loci) and to elucidate their roles in BD, we performed the first functional genomics study of BD. We first systematically identified risk SNPs that disrupted the binding of TFs (these SNPs were referred to as functional SNPs) by integrating chromatin immunoprecipitation sequencing (ChIP-Seq) and position weight matrix (PWM) data. We then conducted a series of experiments (including reporter gene assays, allele-specific expression (ASE) analysis, TF knockdown, and CRISPR/Cas9-mediated genome editing) to validate

the regulatory effects of the identified functional SNPs. We further identified the potential target genes whose expression in the human brain was regulated by the identified functional SNPs using eQTL analysis. Finally, we investigated the function of *PACSI* (a potential target gene regulated by the identified TF binding-disrupting SNPs rs10896081 and rs3862386) and found that overexpression of *PACSI* affected the density of dendritic spines, suggesting the potential mechanism of this gene in BD. Overall, we systematically identified the functional SNPs in the reported risk loci and characterized the regulatory mechanisms of the identified functional SNPs. In addition, our study linked the functional SNPs to their potential target genes, providing a starting point for further functional characterization and development of therapeutic drugs.

## Methods

Information about the reagents and kits used in this study is provided in the Additional file 1.

### Bipolar GWAS used in this study

The genome-wide summary statistics used in this study were obtained from a recent large-scale BD GWAS by Stahl et al. [15]. Stahl et al. first conducted a GWAS on 20,352 BD cases and 31,358 controls (referred to as the discovery stage). Variants with  $P < 1 \times 10^{-4}$  in the discovery stage were then replicated in an additional cohort (9412 cases and 137,760 controls). A total of 30 genome-wide significant (GWS) loci were finally identified by Stahl et al. [15].

### Extraction of SNPs in LD with the index SNPs

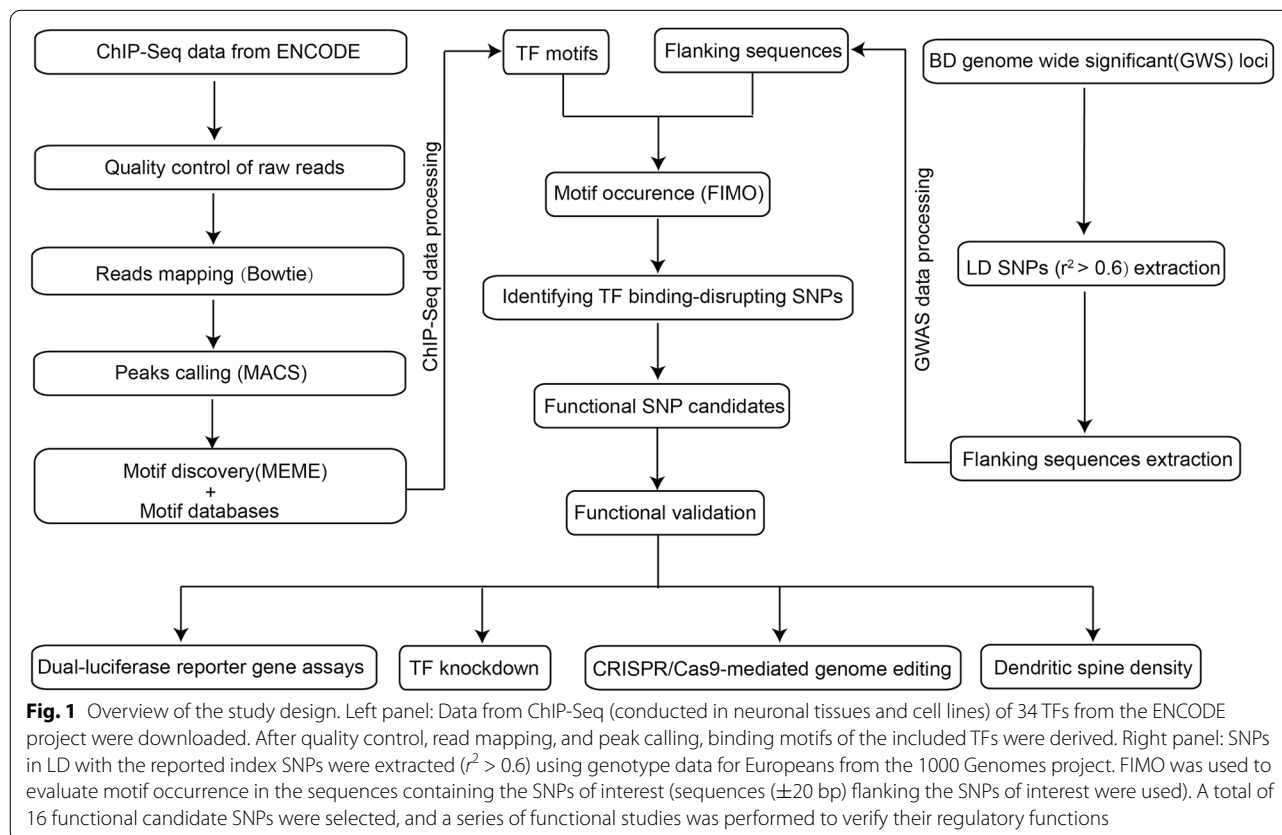
For each risk locus, we extracted SNPs in LD ( $r^2 > 0.6$ ) with the reported index SNPs using genotype data of Europeans from the 1000 Genomes project [18]. Considering that a wide range of LD values ( $r^2$ ) were used across the studies to define whether SNPs of interest were in LD with the reported index SNPs [19–23], we conducted an extensive literature search to select a proper LD threshold in this study. Based on our literature search and the following considerations, we selected the widely accepted LD value ( $r^2 > 0.6$ ) in this study. First,  $r^2 > 0.6$  was widely used to define whether SNPs of interest were in LD with the reported index SNPs [24–37]. Second, we accounted

for both the number of included SNPs and the degree of LD. A higher  $r^2$  (e.g. 0.8) reduces the number of included SNPs and leads to the omission of some potential functional SNPs. Finally, in some cases, the functional SNPs might be in low LD with the reported index SNPs [38–40]. We thus selected the widely used  $r^2$  threshold ( $r^2 > 0.6$ ) in this study. PLINK was used to calculate the LD values and extract the SNPs in LD with the reported index SNPs [41]. In total, 2775 SNPs (including the index SNPs and SNPs in LD with the index SNPs) were extracted (Additional file 2, Table S1).

### Identification of SNPs that affected TF binding

To identify functional (or potential causal) SNPs in the reported GWS loci, we used a functional genomic approach, which has been described in detail in previous studies [42–44]. The flowchart of our functional genomic-based approach includes three major steps (Fig. 1). The first step is derivation of the TF binding motifs. We downloaded raw data for ChIP-Seq of 34 TFs (conducted in brain tissues or neuronal cell lines from the ENCODE project) (<https://www.encodeproject.org/>) [45] and conducted a series of analyses. After cleaning by Btrim (<http://graphics.med.yale.edu/trim/>) [46], the cleaned reads for 30 TFs (4 TFs were excluded because

of low quality) were mapped to the reference genome (hg19) by bowtie (<http://bowtie-bio.sourceforge.net/index.shtml>) [47]. The mapped sam files were then converted into bam files by SAMtools (<http://samtools.sourceforge.net>) [48]. The derived bam files were used for peak calling (by using MACS (<http://liulab.dfci.harvard.edu/MACS/>) [49]). The peaks were sorted, and the sequences of the top 500 ChIP-Seq peaks for each TF were used to derive the binding motifs with the MEME online tool kit (<https://meme-suite.org/meme/tools/meme>) [50]. The derived motifs for each TF were then compared with PWMs (compiled by Whittington et al. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70770>) [44], and the matching motifs were used for further analysis. The second step is extraction of the SNPs in LD with the reported index SNPs. We extracted sequences covering  $\pm 20$  bp around the SNPs of interest (including the index SNPs and SNPs in LD with the index SNPs) (based on the human reference genome (hg19)). For each SNP, two DNA sequences were generated with a difference of only one base at the SNP position (i.e. one sequence contained the reference allele, while the other sequence contained the alternative allele). FIMO (<https://meme-suite.org/meme/tools/fimo>) [51] was used to search for motif occurrence in the DNA sequences (with the  $P$



value threshold set at  $< 0.001$  [42–44]. The final step is identification of the TF binding–disrupting SNPs. We then defined a SNP as a TF binding–disrupting SNP by the following criteria: First, the sequences surrounding the SNP of interest contained a TF binding motif (e.g. a CTCF binding motif) (motif occurrence revealed by FIMO). Second, we checked whether the corresponding TF bound to the sequence containing the SNP of interest in the ChIP-Seq data (from ENCODE). If a SNP met both of the above criteria (i.e. motif occurrence and binding to the corresponding TF (ChIP-Seq peak data)), this SNP was defined as a TF binding–disrupting SNP.

### eQTL analysis

We examined the associations between the identified functional SNPs and gene expression in the human brain by using the original reported eQTL results for 5 eQTL datasets [52–56]. The PsychENCODE project includes human brain tissues from 1866 individuals. eQTL results were downloaded from the PsychENCODE website (<http://resource.psychencode.org/>) [52]. The Common Mind Consortium (CMC) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30272>) [53] eQTL data are based on 209 schizophrenia cases and 206 healthy controls, as well as 52 affective/mood disorder (AFF) cases (tissues from the dorsolateral prefrontal cortex were used for gene expression measurements). The Lieber Institute for Brain Development (LIBD2) (<http://eqtl.brainseq.org/phase2/eqtl/>) [56] eQTL dataset is based on 286 schizophrenia cases and 265 healthy controls (tissues from hippocampus and the dorsolateral prefrontal cortex were used for expression quantification using RNA sequencing). The xQTL (<http://mostafavilab.stat.ubc.ca/xQTLServe/>) [54] dataset contains eQTL data for 494 individuals (tissues from the dorsolateral prefrontal cortex were used for gene expression quantification). The Genotype-Tissue Expression (GTEx) project (<https://gtexportal.org/home/>) [55] contains 49 tissues with eQTL data (sample size ( $N$ ) = 836). Thirteen brain tissues were included in our eQTL analysis. More details about the GTEx project can be found in the original paper and on the GTEx website (<https://gtexportal.org/home/>) [55]. The eQTL results were extracted directly from the original eQTL studies. If the original eQTL results had been subjected to false discovery rate (FDR) correction, only SNP-gene associations with  $FDR < 0.05$  were retained. If the eQTL results did not contain FDR information, SNP-gene associations were adjusted using the Bonferroni correction.

### Allele-specific expression analysis

Allele-specific expression (ASE) analysis is another way to study the regulatory effect of a variant. The

imbalance in expression between two different alleles (maternal and paternal) reflects the potential regulatory effect of a variant. RNA sequencing (RNA-Seq) data can be used to estimate the ASE effect. Suppose that an individual is heterozygous (e.g. with Allele1/Allele2) for a specific SNP. If this SNP is located in a transcribed region, RNA-Seq can determine the number of reads (read counts) containing either Allele1 or Allele2 (quantified as  $Count_{allele1}$  and  $Count_{allele2}$ ). The  $Count_{allele1}/Count_{allele2}$  ratio is compared with  $(Count_{allele1} + Count_{allele2})/2$ :  $(Count_{allele1} + Count_{allele2})/2$  (no ASE) by a binomial test to determine if the number of reads containing Allele1 is significantly different from the number of reads containing Allele2. ASE data in human brain tissues were downloaded from the GTEx project [55] (Version 8). Among the 16 identified functional SNPs, SNPs showing ASE were extracted directly from the GTEx project. More detailed information about ASE analysis in human brain tissues can be found in the original paper and on the GTEx website (<https://gtexportal.org/home/>) [55].

To test whether the identified TF binding–disrupting SNPs showed significant ASE, we performed a Binomial test (using the `binom.test` function implemented in R). The total number of variants reported by GTEx V8 is 46,526,292 [55], and 571,220 variants show significant ASE in brain tissues [55, 57]. We tested whether 4 of the 16 functional SNPs showing ASE were statistically significant by running the command `binom.test(4, 16, 571220/46526292)` implemented in R.  $571220/46526292$  is the probability if a random SNP showing significant ASE (randomly selected from all the variants of GTEx panel), and Binomial test will determine whether 4 of the 16 function SNP showing significant ASE is statistically significant.

### Cell culture

Three cell lines (HEK293T, SH-SY5Y, and U251) were originally obtained from the Kunming Cell Bank at the Kunming Institute of Zoology, Chinese Academy of Sciences. HEK293T and U251 cells were cultured in high-glucose DMEM (Gibco, Cat. No: C11995500BT) supplemented with 10% FBS (Gibco, Cat. No: 10091148) and 1% penicillin and streptomycin (100 U/mL). SH-SY5Y cells were cultured in high-glucose DMEM (Gibco, Cat. No: C12430500BT) supplemented with 10% FBS (Gibco, Cat. No: 10091148), 10 mM sodium pyruvate solution (Gibco, Cat. No: 11360070), 1% penicillin and streptomycin (100 U/mL), and  $1\times$  minimum essential medium nonessential amino acid solution (Gibco, Cat. No: 11140050). All cells were cultured at 37 °C in 5% CO<sub>2</sub>. All cell lines were confirmed to be mycoplasma-free by regular testing by PCR analysis.

### Reporter gene assays

DNA fragments (approximately 500 bp) located in the promoter regions containing different alleles of the candidate functional SNPs were inserted into the pGL4.11[luc2P] vector (which is used to determine promoter activity); alternatively, the enhancer regions containing different alleles of the candidate functional SNPs were inserted into the pGL3-Promoter vector (which is used to determine enhancer activity). The constructs were validated by Sanger sequencing.

We performed reporter gene assays in three cell lines (HEK293T, SH-SY5Y, and U251) as previously described [42, 43]. HEK293T, SH-SY5Y, and U251 cells were plated into 96-well plates at densities of  $3.0 \times 10^4$  cells/well,  $1.0 \times 10^5$  cells/well, and  $1.0 \times 10^4$  cells/well, respectively. After culture for 12 h, the constructed vectors and the internal control vector (pRL-TK, Promega, Cat. No: E2241) were cotransfected into cells using Lipofectamine™ 3000 (Invitrogen, Cat. No: L3000-015). For HEK293T cells, 100 ng of the constructed vectors and 20 ng of the pRL-TK were used. For SH-SY5Y and U251 cells, 150 ng of the constructed vectors and 50 ng of the pRL-TK vector were used. Forty-eight hours post transfection, luciferase activity was measured by a dual-luciferase reporter gene assay system (Promega, Cat. No: E1960) according to the manufacturer's instructions. Differences were calculated with two-tailed Student's *t* test, and the significance threshold was set at  $P < 0.05$ .

### Knockdown of TFs

We used online short hairpin RNA (shRNA) design tools (<http://rnaidesigner.thermofisher.com/rnaiexpress/setOption.do?designOption=shrna&pid=-3105315568901923019>) [58] to design shRNAs targeting CTCF, PBX3, and TAF1. The sequences of the shRNAs are provided in Additional file 1, Table S2. The annealed oligos were cloned into the pLKO.1 vector, and the constructs were validated by Sanger sequencing. Lentiviral particles were obtained by cotransfecting the constructed vectors (10 µg) with the envelope plasmid pMD2.G (2 µg, Addgene, Cat. No: 12259) and the packaging plasmid psPAX2 (5 µg, Addgene, Cat. No: 12260) into HEK293T cells. Forty-eight hours post transfection, the supernatant containing the packaged lentiviral particles were collected for SH-SY5Y cell infection. The cells were then subjected to puromycin (2 µg/mL, Sigma, Cat. No: 540222) treatment for 1 week to select the cells with stable expression of the shRNAs of interest. The TF knockdown efficiency was determined by real-time quantitative PCR (RT-qPCR).

### Deletion of genomic sequences containing the identified functional SNPs by CRISPR/Cas9 genome editing

To evaluate whether the target genes of interest are regulated by the genomic regions containing the candidate functional SNPs, we used CRISPR/Cas9 technology to delete the genomic regions containing the target SNPs. We designed a pair of sgRNAs (sgRNA1 and sgRNA2, respectively) for each target SNP using the CRISPR sgRNA Design Tool (<https://zlab.bio/guidedesign-resources>). The plasmids PX459M and EZ-GuideXH were first linearized with the restriction enzyme BbsI, and sgRNA1 and sgRNA2 were inserted into PX459M and EZ-GuideXH, respectively. After validation by Sanger sequencing, the cassette expressing sgRNA2 from EZ-GuideXH was subcloned into a linearized PX459M plasmid that contained sgRNA1 with the restriction enzymes HindIII and XhoI. The ClonExpress II One Step Cloning Kit (Vazyme, Cat. No: C112-01) was employed to generate the final recombinant plasmid expressing both sgRNA1 and sgRNA2 to perform genome editing in HEK293T cells.

### Real-time quantitative PCR (RT-qPCR) analysis

Total RNA (1 µg) was used as templates for reverse transcription by using the PrimeScript™ RT Kit with gDNA Eraser (Takara, Cat. No: RR047A). The generated cDNA was diluted 1:5 for subsequent RT-qPCR analysis, which was carried out using TB Green™ Premix Ex Taq™ II (Tli RNaseH Plus) (Takara, Cat. No: RR820A) in a QuantStudio™ 12K Flex (Applied Biosystems) instrument or a CFX96 Touch™ Real-Time PCR detection system according to the manufacturers' instructions. *ACTB* was used as the internal control, and the  $2^{-\Delta\Delta C_t}$  method [59] was used to calculate relative gene expression. The significance threshold was set at  $P < 0.05$ , and differences were calculated with two-tailed Student's *t* test. Primer sequences are provided in Additional file 1, Table S3.

### Dendritic spine density analysis

#### Animals

Wild-type C57BL/6J mice were purchased from Shanghai Model Organisms Center (<http://www.modelorg.com>), and the animals were maintained in a quiet, uncrowded temperature-controlled house on a 12-h light/dark cycle (lights on at 08:00 and lights off at 20:00) with ad libitum access to lab chow and water. All experiments were approved by the Animal Ethics Committee of the Kunming Institute of Zoology (License number: SMKX-2021-01-001) and conformed to National Advisory Committee for Laboratory Animal Research guidelines.

### **Culture of mouse cortical neurons**

The pregnant C57BL/6J mice were anaesthetized and euthanized using a CO<sub>2</sub> chamber. Brain tissue was harvested from more than 5 mouse embryos (E16.5–17.5), and the prefrontal cortices were isolated in 1× HBSS. The cortical tissues were digested with papain (Worthington, Cat. No: LS003119) and DNase I (Sigma, Cat. No: D4263-1VL) at 37 °C for 18 min. The digested tissues were then dissociated to obtain single-cell suspensions of neurons. Neurons were seeded into 6-well plates (containing coverslips precoated with poly-D-lysine hydrobromide (Sigma, Cat. No: P6407-5MG; 10 µg/mL)) and cultured in 2 mL of culture medium (neurobasal medium (Gibco, Cat. No: 21103049), 2% B27 (Gibco, Cat. No: 17504044), 1× GlutaMAX<sup>TM</sup>-I (Gibco, Cat. No: 35050061), and 2.5% FBS (Biological Industries, Cat. No: 04-001-1ACS)). After 4 h, the medium was changed to culture medium without FBS (neurobasal medium, 2% B27, and 1× GlutaMAX<sup>TM</sup>-I). Cultures were incubated at 37 °C in a humidified, 5% CO<sub>2</sub> atmosphere for 14 days. Half of the culture medium was refreshed every 7 days.

### **Plasmid transfection and immunofluorescence staining**

The recombinant pCDH constructs for *PACSI1* (or control vector (pCDH-GFP)) and Venus vector were cotransfected into cultured neurons (cultured for 14 days) using Lipofectamine 3000. After 3 days, the neurons were first fixed with 4% paraformaldehyde and 4% sucrose dissolved in PBS at room temperature and were then treated with 0.1% Triton X-100 and 2% goat serum in PBS. The neurons were stained with anti-mCherry (GeneTex, Cat. No: GTX128508) and anti-GFP (Abcam, Cat. No: ab13970) antibodies overnight at 4 °C and were then incubated with the corresponding secondary antibodies for 1 h at room temperature.

### **Image acquisition and dendritic spines analyses**

Analyses of dendritic spine density were carried out as previously described [60–62]. Briefly, images of fixed neurons expressing GFP or both GFP and mCherry were acquired at random using an LSM 880 confocal microscope (Carl Zeiss) by Z-stack image scanning (41 images, 0.25-µm intervals, 1024 × 1024 pixel resolution) with a ×100 objective and 10× digital zoom. The intense expression of GFP encoded by the Venus vector was employed to outline the morphology of neuronal dendritic spines. NeuronStudio [63, 64] was used to analyse secondary or tertiary dendritic spines, including their shape and density. Data obtained from more than 2 dendrites (total length of 60–100 µm per

dendrite) of each neuron were averaged as the result for one neuron to reduce variability. Statistical analysis of the total dendritic spine density between the two groups was performed with two-tailed Student's *t* test. Spine subtype (mushroom, thin, and stubby) densities were analysed using a 2-way ANOVA. All statistical analyses were performed with GraphPad Prism 8, and the significance level was set at 0.05. All assays were performed in at least two independent experiments.

## **Results**

### **Functional genomics identified 16 TF binding–disrupting SNPs in the reported BD risk loci**

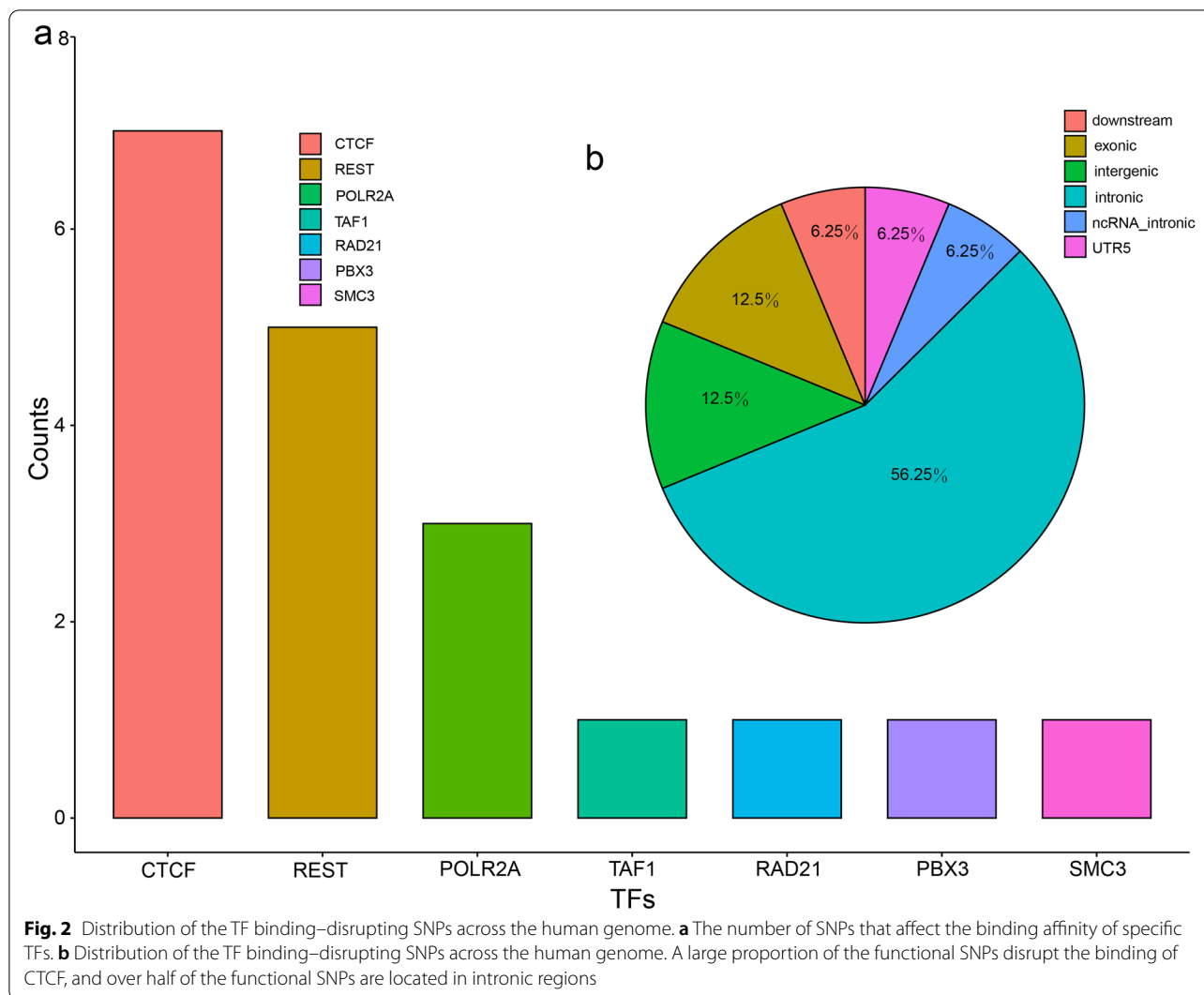
To identify the functional SNPs in the reported risk loci, we utilized a functional genomics approach (Fig. 1) [42, 43]. By integrating ChIP-Seq and PWM data, we identified 16 SNPs (Additional file 1, Table S4) that affected the binding of TFs (these TF binding–disrupting SNPs were called functional SNPs). ANNOVAR annotations [65] showed that most of the SNPs were located in intronic ( $N = 9$ ) and intergenic ( $N = 2$ ) regions (Fig. 2). Of note, among the 16 functional SNPs, 7 affected the binding of CTCF, and 5 affected REST binding (Fig. 2). In addition, 3 SNPs affected the binding of two or more TFs: rs2027349 (affected CTCF/TAF1 binding), rs3862386 (affected CTCF/REST binding), and rs228769 (affected CTCF/SMC3 binding) (Additional file 1, Table S4). These results identified functional SNPs in the reported risk loci, suggesting that these functional SNPs may exert their effects on BD by regulating gene expression.

### **Reporter gene assays validated the regulatory effects of the identified functional SNPs**

Our functional genomic analysis identified 16 TF binding–disrupting SNPs in the reported BD risk loci. To verify the effects of the identified functional SNPs, we carried out dual-luciferase reporter gene assays in HEK293T, SH-SY5Y, and U251 cells. Among the 16 TF binding–disrupting SNPs, all exhibited regulatory effects (i.e. different alleles of these SNPs significantly affected luciferase activity) in at least one tested cell lines (Additional file 1, Table S5). Of note, 11 SNPs exhibited regulatory effects in all three cell lines (Figs. 3, 4, 5, and 6), and 3 SNPs showed regulatory effects in both SH-SY5Y and U251 cells (Additional file 1, Figure S1). These results validated the regulatory effects of these identified functional SNPs.

### **Disruption of PBX3 binding by rs10896081**

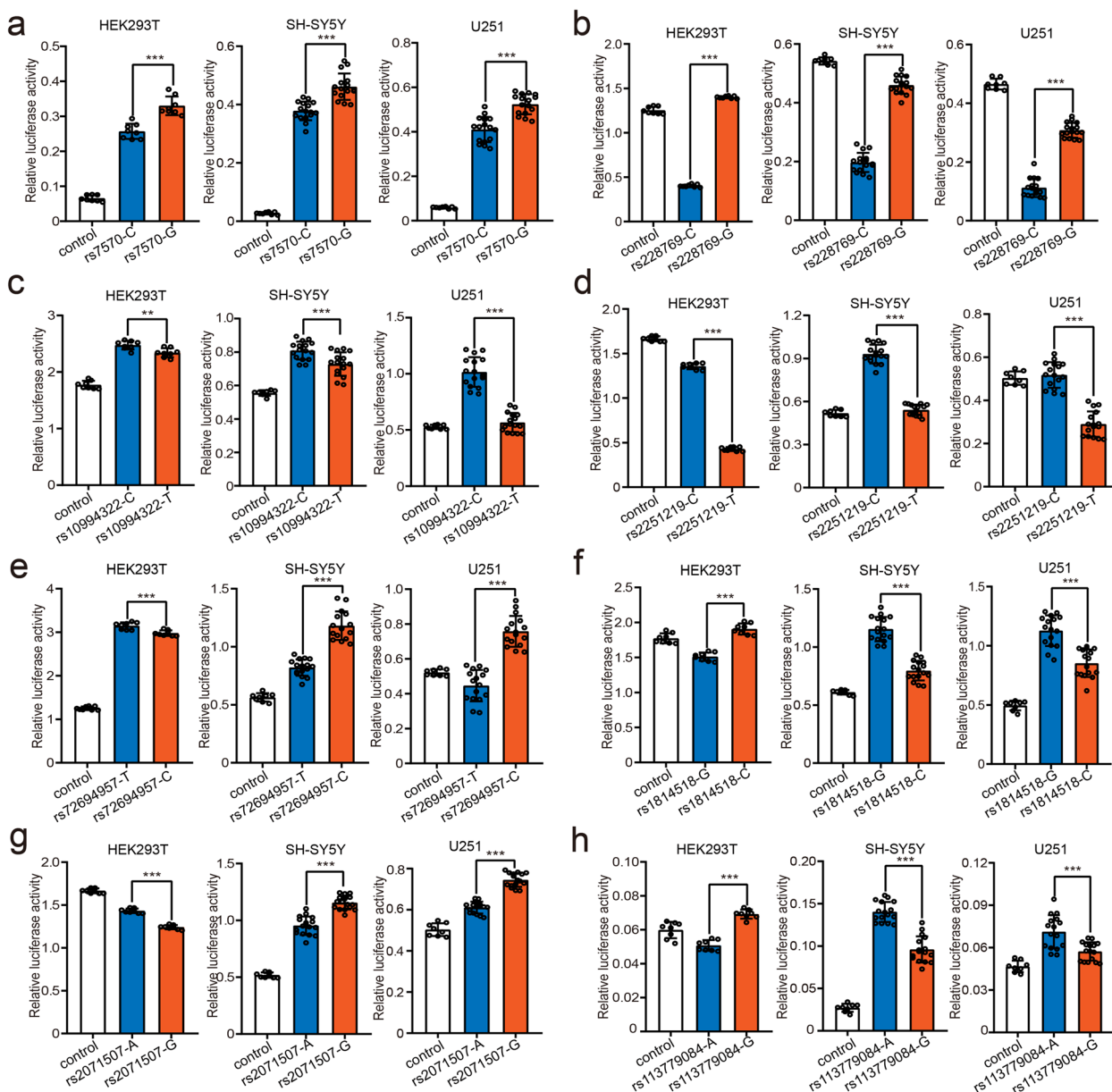
To further investigate the regulatory mechanisms of the functional SNPs, we focused on SNP rs10896081 (located at 11q13.2), which disrupted the binding of



the TF PBX3 (Fig. 4a). The ChIP-Seq data indicated that PBX3 can bind to the genomic region containing rs10896081 (Fig. 4b). The DNase-Seq data showed that rs10896081 is located in a genomic region with active transcription in brain tissues (or neuronal cell lines) (Fig. 4b). The histone modification data revealed that rs10896081 is located in a chromosomal region marked with H3K27ac (a marker for active enhancers [66, 67]) (Fig. 4b). These data indicated that rs10896081 is located in an active regulatory element. To test whether rs10896081 has a functional consequence, we performed reporter gene assays. We found that the T allele of rs10896081 produced significantly higher activity than the A allele in all three tested cell lines (Fig. 4c), supporting the regulatory effect of rs10896081. We then conducted eQTL analysis to identify genes whose expression levels in the human brain are associated with rs10896081. Four genes (*PACSI*, *RP11-755F10.1*,

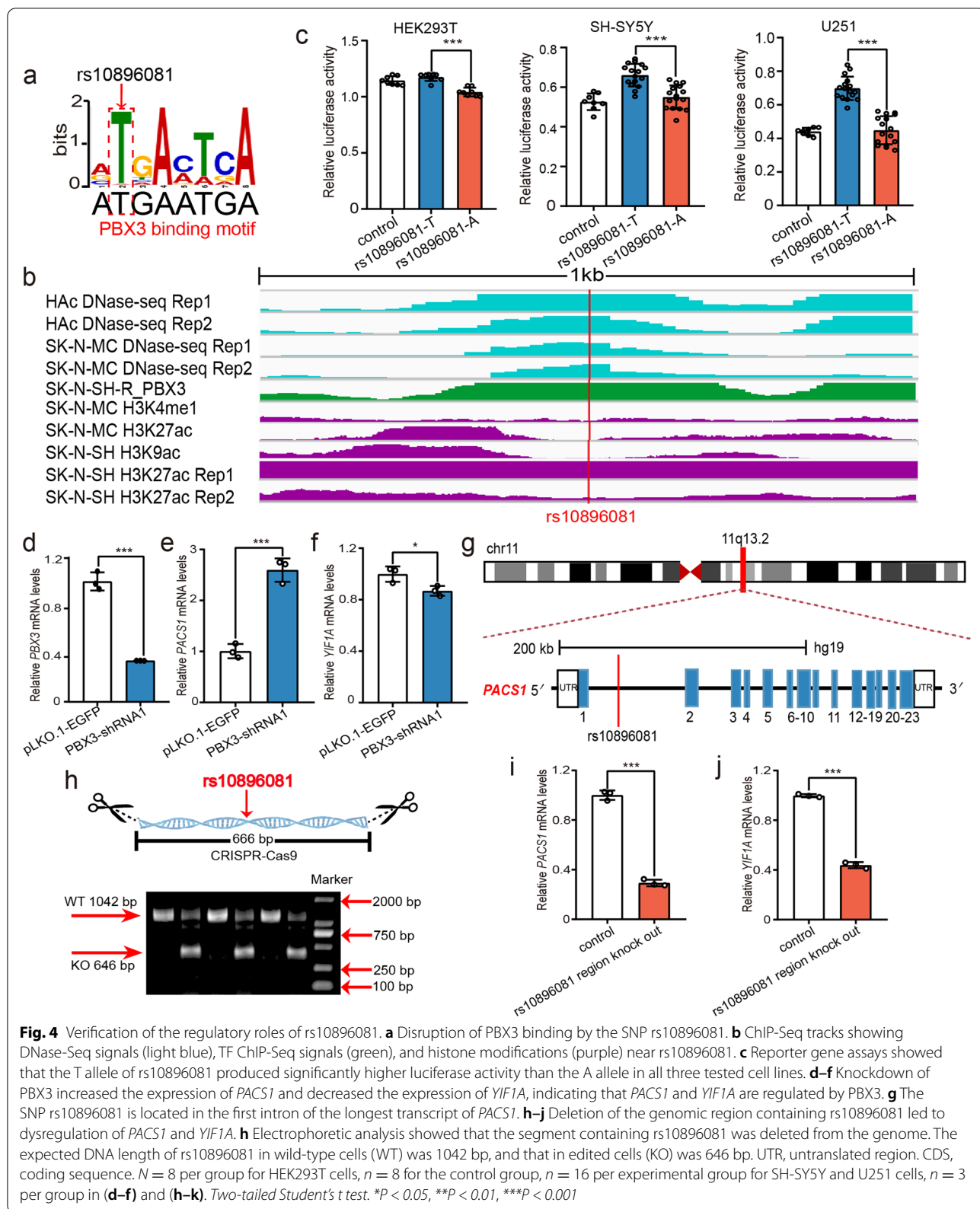
*RAB1B*, and *YIF1A*) showed the most significant associations with rs10896081 in the human brain (Additional file 2, Table S6), suggesting that these genes were potential target genes of rs10896081.

The finding that rs10896081 disrupts the binding of PBX3 implies that rs10896081 regulates its eQTL genes by affecting PBX3 binding. To further validate whether *PACSI*, *RP11-755F10.1*, *RAB1B*, and *YIF1A* are regulated by PBX3, we repressed *PBX3* expression in SH-SY5Y cells. *PBX3* knockdown resulted in significant dysregulation of *PACSI* and *YIF1A* (Fig. 4d–f), indicating regulatory effects of PBX3 on these two genes. The expression of *RP11-755F10.1* was not examined, as it is a pseudogene. In addition, *RAB1B* did not show a significant expression change in cells with *PBX3* repression. As rs10896081 is located in the first intron of the longest transcript of *PACSI* (Fig. 4g), we further investigated the regulatory effect of the genomic region containing

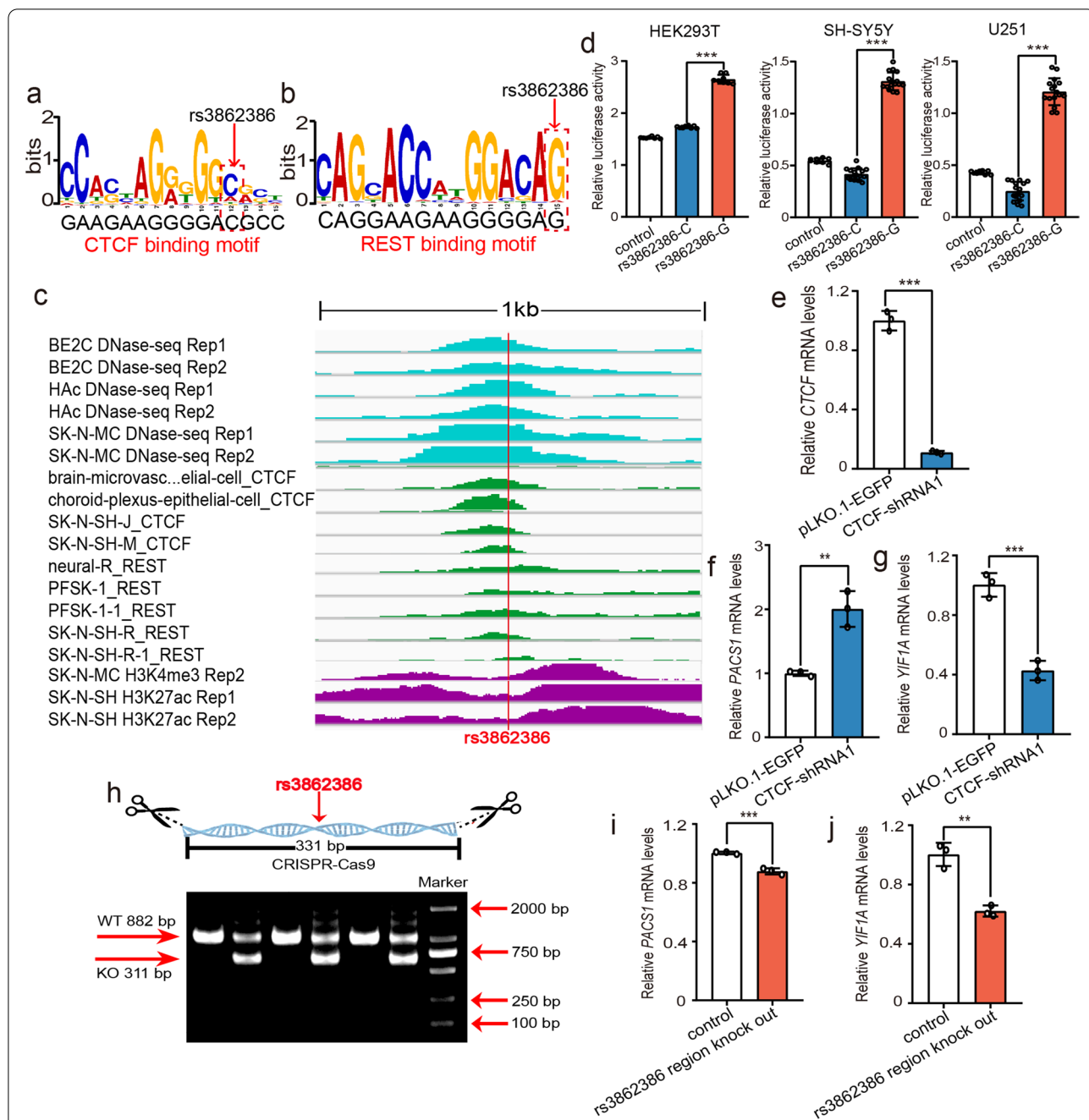


**Fig. 3** Validation of the regulatory effect of the TF binding–disrupting SNPs. **a** The luciferase expression of the constructs carrying the G allele of rs7570 was significantly higher than that of constructs carrying the C allele in all three tested cell lines. **b** The G allele of rs228769 showed significantly higher luciferase activity than the C allele in all three tested cell lines. **c** The constructs containing the C allele of rs10994322 produced significantly higher luciferase activity than the constructs containing the T allele in all three tested cell lines. **d** The constructs carrying the C allele of rs2251219 exhibited significantly higher luciferase activity than the constructs carrying the T allele in all three tested cell lines. Notably, this result was inconsistent with the previously published study by Yang et al. [61], likely because of the different lengths (in this study, 513 bp; in Yang et al. study, 435 bp) and directions (in this study, 5' to 3'; in yang et al. study, 3' to 5') of the DNA fragments containing rs2251219 in the pGL3-Promoter vector between the studies. **e** The reporter vectors containing the C allele of rs72694957 showed significantly higher luciferase activity than those containing the T allele in SH-SY5Y and U251 cells. **f** The reporter vectors containing the G allele of rs1814518 displayed significantly higher luciferase activity than those containing the C allele in SH-SY5Y and U251 cells. **g** The luciferase expression of the vectors containing the G allele of rs2071507 was significantly higher than that of the vectors containing the A allele in SH-SY5Y and U251 cells. **h** The constructs carrying the A allele of rs113779084 produced significantly higher luciferase activity than those carrying the G allele in SH-SY5Y and U251 cells.  $N = 8$  per group for HEK293T cells,  $n = 8$  for the control group, and  $n = 16$  per experimental group for SH-SY5Y cells and U251 cells. Two-tailed Student's *t* test. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$





**Fig. 4** Verification of the regulatory roles of rs10896081. **a** Disruption of PBX3 binding by the SNP rs10896081. **b** ChIP-Seq tracks showing DNase-Seq signals (light blue), TF ChIP-Seq signals (green), and histone modifications (purple) near rs10896081. **c** Reporter gene assays showed that the T allele of rs10896081 produced significantly higher luciferase activity than the A allele in all three tested cell lines. **d-f** Knockdown of PBX3 increased the expression of *PACS1* and decreased the expression of *YIF1A*, indicating that *PACS1* and *YIF1A* are regulated by PBX3. **g** The SNP rs10896081 is located in the first intron of the longest transcript of *PACS1*. **h-j** Deletion of the genomic region containing rs10896081 led to dysregulation of *PACS1* and *YIF1A*. **h** Electrophoretic analysis showed that the segment containing rs10896081 was deleted from the genome. The expected DNA length of rs10896081 in wild-type cells (WT) was 1042 bp, and that in edited cells (KO) was 646 bp. UTR, untranslated region. CDS, coding sequence.  $N = 8$  per group for HEK293T cells,  $n = 8$  for the control group,  $n = 16$  per experimental group for SH-SY5Y and U251 cells,  $n = 3$  per group in (**d-f**) and (**h-k**). Two-tailed Student's *t* test. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$



**Fig. 5** Confirmation of the regulatory effect of rs3862386. **a, b** The SNP rs3862386 disrupts CTCF and REST binding. **c** The SNP rs3862386 is located in a genomic region with DNase-Seq, ChIP-Seq, and histone modification signals, indicating that it is located in a genomic region with active transcription in neuronal cells. **d** Reporter gene assays showed that the G allele of rs3862386 produced significantly higher luciferase activity than the C allele in all three tested cell lines. **e-g** The expression of *PACS1* and *Y1F1A* was significantly altered by CTCF knockdown. **h-j** A 331-bp genomic sequence containing rs3862386 was deleted by CRISPR/Cas9-mediated genome editing. Deletion of rs3862386 resulted in altered expression of *PACS1* and *Y1F1A*. **h** Genomic PCR/electrophoresis results showed deletion of the genomic region containing rs3862386. WT, genomic DNA containing rs3862386 in wild-type cells (882 bp). KO, genomic DNA containing rs3862386 in edited cells (311 bp). *N* = 8 per group for HEK293T cells, *n* = 8 for the control group, *n* = 16 per experimental group for SH-SY5Y and U251 cells, *n* = 3 per group in (**e-h**) and (**j-l**). Two-tailed Student's *t* test. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001

rs10896081 on *PACSI* and *YIF1A* using CRISPR/Cas9. Deletion of the genomic sequence (666 bp) containing rs10896081 (Fig. 4h; Additional file 1, Figure S2) led to significant decreases in the expression of *PACSI* and *YIF1A* (Fig. 4i, j). Taken together, these results suggested that rs10896081 regulates the expression of *PACSI* and *YIF1A* by interacting with PBX3.

#### Regulatory mechanisms of rs3862386

We also investigated the regulatory mechanism of rs3862386, a SNP that affects the binding of CTCF and REST (Fig. 5a, b). SNP rs3862386 is located in a genomic region with strong ChIP-Seq, DNase-Seq, and histone modification signals (Fig. 5c), indicating that rs3862386 lies in a regulatory element to which the TFs CTCF and REST bind. Reporter gene assays showed that the G allele (protective allele) of rs3862386 was associated with higher luciferase activity than the C allele in all three tested cell lines, confirming the regulatory effect of rs3862386. Interestingly, rs3862386 is in strong LD ( $r^2 = 0.97$ ) with rs10896081 (a SNP that disrupts the binding of PBX3). To explore the potential target genes regulated by rs3862386, we conducted eQTL analysis and found that rs3862386 was associated with the expression of *PACSI*, *RP11-755F10.1*, *RAB1B*, and *YIF1A* in the human brain (Additional file 2, Table S6). We thus further investigated whether rs3862386 and its binding TFs (CTCF and REST) regulate the expression of the four eQTL genes of rs3862386. Knockdown of CTCF in SH-SY5Y cells resulted in significant upregulation of *PACSI* and downregulation of *YIF1A* (Fig. 5e–g), suggesting that the expression of these genes is regulated by CTCF. However, no alteration in *RAB1B* expression was detected. The expression of *RP11-755F10.1* was not determined because it is a proposed pseudogene. We then further analysed whether rs3862386 regulates the expression of *PACSI* and *YIF1A*. Deletion of the genomic sequence (331 bp) containing rs3862386 (Fig. 5h; Additional file 1, Figure S3) led to dysregulation of *PACSI* and *YIF1A* (Fig. 5i, j), indicating the regulatory effect of rs3862386 on *PACSI* and *YIF1A*. These data suggested that rs3862386 may confer risk for BD by modulating *PACSI* and *YIF1A* expression.

#### Disruption of CTCF and TAF1 binding by rs2027349

We characterized rs2027349, a SNP that affects the binding of the TFs CTCF and TAF1 (Fig. 6a, b). The ChIP-Seq data revealed that the TFs CTCF and TAF1 can bind to the genomic sequence containing rs2027349 in cell lines from the human brain, and the DNase-Seq and histone modification data showed that rs2027349 is located in an actively transcribed region (Fig. 6c). Reporter gene assays further validated the regulatory role of rs2027349. The G allele of rs2027349 produced significantly higher luciferase activity than the A allele in all three tested cell lines (Fig. 6d), indicating the regulatory function of rs2027349. eQTL analysis showed that rs2027349 was associated with *ANP32E*, *TARS2*, *RPRD2*, and *VPS45* expression in the human brain (uncorrected  $P < 0.01$  in at least one eQTL dataset) (Additional file 2, Table S6), suggesting that rs2027349 regulates these genes.

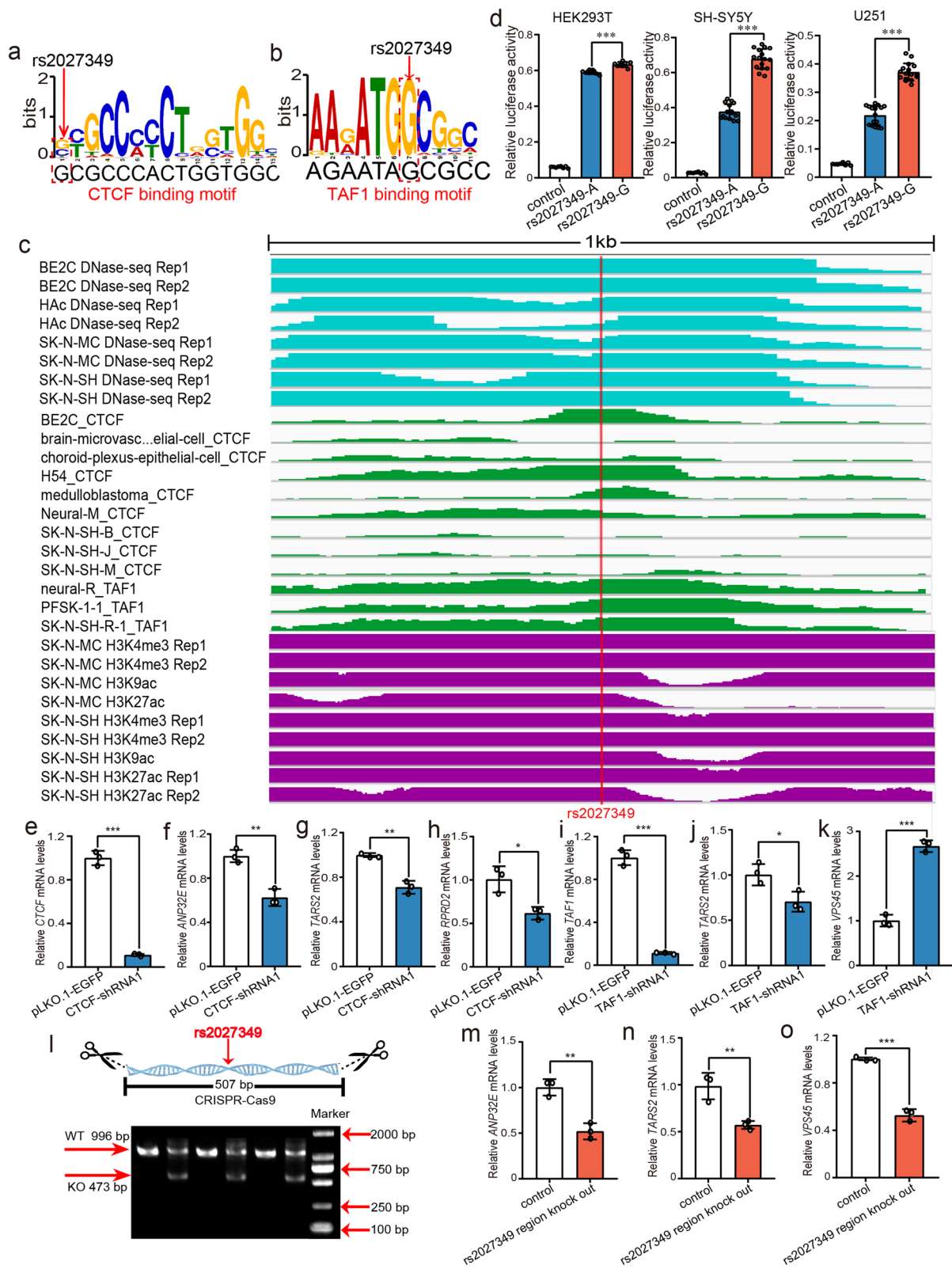
To further explore whether rs2027349 regulates its eQTL genes via interactions with CTCF and TAF1, we knocked down CTCF and TAF1. CTCF knockdown resulted in significant alterations in *ANP32E*, *TARS2*, and *RPRD2* expression, but *VPS45* expression did not change in cells with CTCF repression (data not shown). In addition, TAF1 knockdown led to dysregulation of *TARS2* and *VPS45* (Fig. 6e–k), indicating the regulatory effect of CTCF and TAF1 on these genes. *ANP32E* and *RPRD2* did not show expression changes in TAF1 knockdown cells (data not shown). Finally, rs2027349 deletion (507 bp) led to dysregulation of *ANP32E*, *TARS2*, and *VPS45* (Fig. 6l–o; Additional file 1, Figure S4), supporting the hypothesis that rs2027349 regulates its eQTL genes by interacting with the TFs CTCF and TAF1.

#### eQTL analysis identified the potential target genes regulated by the identified TF binding–disrupting SNPs

To further identify the potential target genes regulated by the identified functional SNPs, we used five human brain eQTL datasets. Among the 16 TF binding–disrupting SNPs, 14 were associated with gene expression (uncorrected  $P < 0.01$ ) in at least one brain eQTL dataset (Additional file 2, Table S6), 12 were significantly correlated with gene expression in at least

(See figure on next page.)

**Fig. 6** Validation of the regulatory effect of rs2027349 by reporter gene assays, TF knockdown and CRISPR/Cas9-mediated genome editing. **a, b** The SNP rs2027349 disrupts CTCF and TAF1 binding. **c** The 1 kb sequence near the SNP rs2027349 is marked with a strong DNase-Seq (light blue), TF ChIP-Seq (green), and histone modification (purple) signals. **d** Reporter gene assays showed that the G allele of rs2027349 produced significantly higher luciferase activity than the A allele in all three tested cell lines. **e–k** CTCF knockdown led to downregulation of *ANP32E*, *TARS2*, and *RPRD2*. However, knockdown of TAF1 resulted in downregulation of *TARS2* and upregulation of *VPS45*. **l–o** A genomic sequence containing rs2027349 was deleted by CRISPR/Cas9-mediated genome editing, which resulted in downregulation of *ANP32E*, *TARS2*, and *VPS45* expression. **l** Electrophoresis showed that the genomic region containing rs2027349 was deleted by the sgRNAs. WT, genomic DNA containing rs2027349 (996 bp) in wild-type cells. KO, genomic DNA containing rs2027349 (473 bp) in edited cells.  $N = 8$  per group for HEK293T cells,  $n = 8$  for the control group,  $n = 16$  per experimental group for SH-SY5Y and U251 cells,  $n = 3$  for each group in (**e–k**) and (**l–o**). Two-tailed Student's *t* test. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$



**Fig. 6** (See legend on previous page.)

**Table 1** Association significance between the TF binding-disrupting SNPs and gene expression in the human brain tissues(at least four brain eQTL datasets)

SNP	Gene	CMC <sup>a</sup>		LIBD2(DLPFC) <sup>a</sup>		LIBD2(Hippo) <sup>a</sup>		xQTL <sup>a</sup>		Brain(GTEX) <sup>c</sup>	
		P value	FDR	P value	FDR	P value	FDR	P value	FDR <sup>b</sup>	P value	P value
rs2071507	GLT1D1	2.02E-08	2.80E-06	2.83E-06	1.86E-04	NA	NA	1.30E-03	4.69E-02	0.000030(Brain - Nucleus accumbens (basal ganglia))	0.000005(Brain - Cerebellar Hemisphere);
rs2071507	NEK4	1.25E-12	3.32E-10	4.63E-08	4.39E-06	2.93E-06	2.76E-04	2.68E-04	1.28E-02	7.0e-7(Brain - Cortex);0.000044( Brain - Frontal Cortex (BA9));	9.9e-7(Brain - Cerebellar Hemisphere);0.0000062(Brain - Cerebellum)
rs3862386	PACS1	2.30E-12	6.51E-10	9.28E-09	9.95E-07	NA	NA	3.69E-10	5.78E-08	0.000010(Brain - Cortex);0.0000044(Brain - Frontal Cortex (BA9));	3.9e-7(Brain - Cortex);0.0000068(Brain - Frontal Cortex (BA9));
rs2251219	NEK4	5.00E-16	2.01E-13	3.71E-17	1.15E-14	8.40E-12	2.07E-09	1.14E-06	1.00E-04	0.0000016(Brain - Hypothalamus)	0.0000004(Brain - Frontal Cortex (BA9));
rs6591201	PACS1	5.69E-06	5.67E-04	7.98E-05	3.56E-03	1.93E-05	1.46E-03	4.51E-09	6.05E-07	0.0000004(Brain - Cortex);0.0000030(Brain - Frontal Cortex (BA9));	0.0000004(Brain - Cortex);0.0000030(Brain - Frontal Cortex (BA9));
rs2270448	PACS1	3.66E-09	6.71E-07	1.54E-06	1.08E-04	NA	NA	5.76E-12	1.14E-09	0.0000004(Brain - Cortex);0.0000030(Brain - Frontal Cortex (BA9));	0.0000004(Brain - Cortex);0.0000030(Brain - Frontal Cortex (BA9));
rs10896081	PACS1	2.21E-12	6.29E-10	9.13E-08	8.21E-06	NA	NA	6.60E-10	9.97E-08	0.0000004(Brain - Cortex);0.0000030(Brain - Frontal Cortex (BA9));	0.0000004(Brain - Cortex);0.0000030(Brain - Frontal Cortex (BA9));

<sup>a</sup> Brain tissues from the dorsolateral prefrontal cortex (DLPFC) were used in the CommonMind Consortium (CMC) (N=467), the xQTL map of the human brains (xQTL) (N=494) and the Lieber Institute for Brain Development (LIBD2) brain eQTL (N=551). <sup>b</sup>The FDR were calculate by R function p.adjust(). <sup>c</sup>FDR value is not available for GTEX eQTL data, but GTEX provided genome-wide empirical P-value threshold for each gene that calculated under FDR=0.05, all the GTEX eQTL P values listed here had passed the P-value threshold (FDR < 0.05)

two brain eQTL datasets (Additional file 2, Table S7), 9 exhibited significant associations with gene expression in at least three brain eQTL datasets (Additional file 2, Table S8), and 6 were significantly associated with gene expression in at least four brain eQTL datasets (Table 1). Notably, 3 SNPs showed significant associations with gene expression in all five brain eQTL datasets (Table 1), suggesting that these SNPs may regulate the expression of their target genes. Considering these results collectively, we prioritized the potential target genes regulated by the identified functional SNPs.

#### ASE analysis supported the regulatory effects of the identified functional SNPs

To further explore the regulatory effects of the TF binding–disrupting SNPs, we used ASE data from the GTEx project (including only brain tissues). The results showed that 4 of the 16 TF binding–disrupting SNPs

exhibited ASE in the human brain. The binomial test indicated that 4 out of the 16 SNPs showed ASE was statistically significant compared with the proportion of SNPs showing ASE in the GTEx project ( $P = 3.67 \times 10^{-5}$ , Additional file 2, Table S9), indicating that the TF binding–disrupting SNPs are more likely to show ASE. The four ASE SNPs are rs2027349 (Fig. 6), rs2251219 (Fig. 3d), rs1814518 (Fig. 3f), and rs2270448 (Additional file 1, Figure S1a). These ASE results provided further support for the functionality of the TF binding–disrupting SNPs.

#### Differential expression analysis of genes (eQTL genes) whose expression was associated with the identified functional SNPs

To further verify whether the identified functional SNPs may confer risk for BD by regulating the expression of their target genes, we examined the expression levels of

**Table 2** Differentially expression analysis of genes (eQTL genes) associated with the identified functional SNPs

Ensembl_gene_id	Gene_name	P.value	FDR	Log <sub>2</sub> FC
ENSG0000010322	NISCH	0.000652673	0.025719695	0.053824812
ENSG00000087365	SF3B2	0.025271057	0.191498291	0.034375659
ENSG00000092529	CAPN3	0.025012225	0.190376571	-0.125577335
ENSG00000099785	MARCHF2	0.045287401	0.257423271	0.038582022
ENSG00000105708	ZNF14	0.000256993	0.015221847	0.069962199
ENSG00000117791	MTARC2	0.001085525	0.033910493	0.075874475
ENSG00000160161	CILP2	0.000082	0.007972009	0.220446629
ENSG00000163938	GNL3	0.02844909	0.203075523	0.033097846
ENSG00000166136	NDUFB8	0.047600309	0.263534942	0.037699559
ENSG00000167491	GATAD2A	0.028028595	0.201830947	0.041306987
ENSG00000173599	PC	0.020481736	0.171492954	0.044787537
ENSG00000174791	RIN1	0.006004402	0.086305322	0.067733795
ENSG00000175115	PACS1	0.004206371	0.071508307	0.044807186
ENSG00000180071	ANKRD18A	0.01185869	0.127555154	0.131432268
ENSG00000180979	LRRC57	0.048150452	0.264937063	0.035979853
ENSG00000181638	ZFP41	0.023891485	0.18596537	-0.045816011
ENSG00000187664	HAPLN4	0.035640292	0.227131948	0.098804717
ENSG00000189157	FAM47E	0.012030774	0.128710189	0.062403408
ENSG00000198106	SNX29P2	0.001497159	0.040607774	0.105070975
ENSG00000213533	TMEM110	0.00058669	0.024231067	-0.067704389
ENSG00000215256	DHRS4-AS1	0.014319833	0.141778997	0.095963915
ENSG00000240747	KRBOX1	0.000122545	0.009994372	0.234360192
ENSG00000251867	Y_RNA	0.041863676	0.246888247	0.069471479
ENSG00000254510	RP11-867G23.10	0.000869737	0.029881244	-0.291953622
ENSG00000272414	FAM47E	0.005317918	0.08078778	0.102011622
ENSG00000272573	MUSTN1	0.025553164	0.192617262	-0.266328945
ENSG00000273045	C2ORF15	0.022486423	0.179919816	0.067969908
ENSG00000273170	ANKRD18A	0.017391347	0.156333195	0.161747281
ENSG00000273173	SNURF	0.00818448	0.10344798	0.060262523
ENSG00000273291	KRBOX1	0.01933284	0.165860839	0.157290642

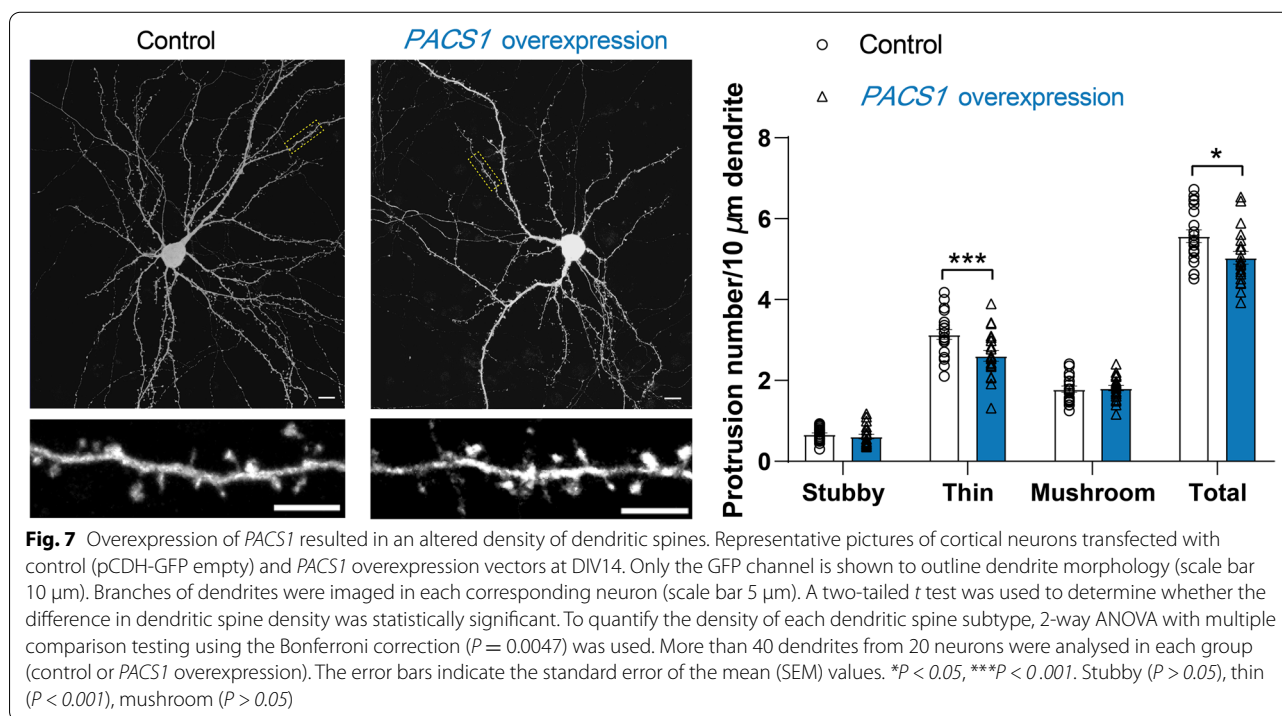
target genes (genes whose expression levels were associated with the functional SNPs) in the brains of BD cases and controls using the transcriptome data from PsychENCODE [68]. Expression analysis showed that 30 target genes were differentially expressed (uncorrected  $P < 0.05$ ) in the brains of BD cases compared with controls (Table 2). Of note, the expression of 8 genes (*NISCH*, *ZNF14*, *MTARC2*, *CILP2*, *SNX29P2*, *TMEM110*, *KRBOX1*, and *RP11-867G23.10*) was significantly dysregulated in BD cases compared with controls (FDR  $< 0.05$ ) (Table 2). These results provided further evidence to support the hypothesis that the identified functional SNPs may contribute to the risk of BD by regulating their target genes.

#### PACSI1 overexpression affected dendritic spine density

To further investigate the potential role of the target genes (those regulated by the identified functional SNPs) in BD, we selected *PACSI1* for further functional characterization. The expression of *PACSI1* was regulated by the TF binding–disrupting SNPs rs10896081 (Fig. 4) and rs3862386 (Fig. 5). In addition, eQTL analysis indicated that *PACSI1* expression was associated with rs10896081 and rs3862386 (Additional file 2, Table S6), suggesting that these two functional SNPs may confer risk for BD by regulating *PACSI1* expression. Notably, expression analysis showed a trend of significant upregulation of *PACSI1* in BD cases compared with controls [68] ( $P = 4.21 \times$

$10^{-3}$ , FDR = 0.072) (Table 2). These convergent and consistent lines of evidence suggest that the functional SNPs rs10896081 and rs3862386 might confer BD risk by regulating *PACSI1* expression.

Accumulating data suggest that dysfunction of dendritic spines (e.g. altered density) may have a pivotal role in BD [17, 69–72]. We thus mimicked the effect of *PACSI1* upregulation on dendritic spine density. To gain insights into the function of *PACSI1* in vitro-cultured primary mouse neurons, we cotransfected the plasmid encoding *PACSI1* (or the control vector) with the Venus plasmid into mouse cortical neurons (day in vitro (DIV) 14). Notably, we observed a significantly decreased total spine density after overexpression of *PACSI1* (control,  $5.568 \pm 0.691$  spines per  $10 \mu\text{m}$ ; *PACSI1* overexpression,  $5.034 \pm 0.691$  spines per  $10 \mu\text{m}$ ; Fig. 7). We further assessed the effects of *PACSI1* overexpression on dendritic spines. Neurons transfected with *PACSI1* showed a selective decrease in the density of immature thin spines with elongated necks and small heads (control,  $3.135 \pm 0.536$  spines per  $10 \mu\text{m}$ ; *PACSI1* overexpression,  $2.610 \pm 0.599$  spines per  $10 \mu\text{m}$ ; Fig. 7). However, the densities of mushroom and stubby spines were not changed. These results indicate the important role of *PACSI1* in mediating the morphogenesis of dendritic spines, suggesting that the identified functional variants rs10896081 and rs3862386 might confer BD risk by modulating *PACSI1* expression.



## Discussion

Since the first report of a BD GWAS in 2007 [73], many BD risk loci have been identified through several larger GWASs in the past decade [13, 15, 74–79]. However, due to the complicated LD and the complexity of gene regulation, identifying causal risk variants in the reported risk loci and elucidating the molecular mechanisms of these causal risk variants in the pathophysiology of BD remain major challenges in the post-GWAS era. In this study, we systematically characterized the regulatory mechanisms of BD risk variants using a functional genomic approach. We identified 16 SNPs (from a total of 2775 SNPs) that disrupted the binding of 7 TFs, and we validated the functional consequences of these identified SNPs through a series of assays, including reporter gene assays, ASE analysis, TF knockdown, and CRISPR/Cas9-mediated genome editing. By combining these approaches with eQTL analysis, we further identified potential target genes regulated by these TF binding-disrupting SNPs. Of note, we showed dysregulation of some target genes (regulated by the identified functional SNPs) in BD cases compared with controls. Finally, we investigated the potential role of *PACSI* (regulated by rs10896081 and rs3862386) in BD pathogenesis.

We noted that approximately 43% (7/16) of the TF binding-disrupting SNPs were located in the CTCF binding motif, implicating that altered CTCF binding may be a common mechanism of BD risk variants. Considering the important role of CTCF in regulating chromosomal conformation [80], our data also suggest that a large proportion of BD risk variants may exert their biological effects by regulating the expression of distal genes. In addition, approximately 56% (9/16) of the TF binding-disrupting SNPs were located in intronic regions, demonstrating the important roles of genetic variants in intronic regions in the regulation of BD risk genes.

The ASE results (Additional file 2, Table S9) provide further evidence indicating that our TF binding-disrupting SNPs are regulatory variants. For example, we showed that rs2027349 (Fig. 6) is a regulatory SNP that affects the binding affinity of CTCF and TAF1. ASE analysis also indicated that rs2027349 showed significant ASE in the human brain (data from GTEx V8, Additional file 2, Table S9), further supporting the regulatory role of rs2027349 in neuronal tissues.

We showed that *PACSI* may have a role in BD pathogenesis. Previous studies have shown that dysfunction of dendritic spines might have a role in BD [69]. *PACSI* (phosphofurin acidic cluster sorting protein 1) encodes a trafficking protein that plays a role in the localization of *trans*-Golgi network (TGN) membrane proteins [81]. *PACSI* has been reported to be associated with many diseases, including acquired immune deficiency syndrome

(AIDS) [82] and Alzheimer's disease [83]. It can induce internalization of MHC-I by interacting with the HIV-1 protein Nef, resulting in reduced immune recognition of infected cells [82]. *PACSI* is also involved in the transport of amyloid precursor protein and enhances the formation of brain plaques in Alzheimer's disease [83]. Mutations in *PACSI* cause a defect in cranial neural crest migration, which leads to intellectual disability [84]. In addition, *PACSI* may play a role in cervical cancer [85]. Although *PACSI* plays an important role in protein transport and is associated with a variety of diseases, the role of this gene in BD is still largely unknown. In this study, we showed that *PACSI* overexpression in mouse primary cortical neurons resulted in an altered density of thin dendritic spines, indicating that *PACSI* may confer risk for BD by affecting the function of dendritic spines.

We noted that some loci contained several TF binding-disrupting SNPs. In our opinion, such TF binding-disrupting SNPs are meaningful for disease susceptibility. First, in our study, we showed that four functional SNPs (rs3862386 ( $r^2 = 0.99$ ), rs10896081 ( $r^2 = 0.99$ ), rs2270448 ( $r^2 = 0.81$ ), and rs6591201 ( $r^2 = 0.73$ )) are in LD with the index SNP rs10896090. As shown in Figs. 4 and 5, the results of reporter assays, TF knockdown, and CRISPR/Cas9-mediated genome editing suggested that rs3862386 and rs10896081 might confer risk for BD by modulating *PACSI* and *YIF1A* expression. In addition, we found that different alleles of rs2270448 and rs6591201 resulted in significant differences in luciferase activity in SH-SY5Y and U251 cells (Additional file 1, Figure S1), suggesting that rs3862386 and rs10896081 are functional variants. In fact, many studies have reported that several functional SNPs in a single risk locus might act synergistically or independently to contribute to disease susceptibility. For example, French et al. showed that two functional SNPs (rs78540526 and rs554219) located in enhancer elements conferred risk for breast cancer through regulating the *CCND1* gene [86]. Chatterjee et al. showed that several regulatory variants in enhancer elements conferred risk for Hirschsprung disease by affecting *RET* expression [87]. Shidal et al. [88] showed that the functional variants rs35418111 and rs2078203 (in LD with the index variant in the 21q22.3 risk locus) might be involved in the occurrence of breast cancer by regulating the expression of *YBEY*. In addition to these reports, other studies also have revealed that several functional variants in a specific risk locus contributed to disease susceptibility by modulating the same risk gene [89, 90]. These results suggest that some loci harbour several functional SNPs to regulate the expression of effector risk genes.

There are a few limitations of this study that should be noted. First, only a limited number of TFs (i.e. 30) were included in this study. Considering that the number of



TFs in the human genome is approximately 1600 [91], we could not identify the risk SNPs that disrupt the binding of other TFs not included in this study. Second, we only analysed single-nucleotide polymorphisms. However, other types of genetic variations (such as copy number variations (CNVs), chromosomal structural variants, rare mutations, and de novo mutations) may also have a pivotal role in BD. These types of genetic variations were not included in our study, and further work is needed to explain the importance of other types of genetic variations in BD. Third, during execution of this study, a larger GWAS on BD was published [16]. The new risk loci identified in this study were not included in our study. Fourth, overexpression of *PACSI* in cultured mouse primary cortical neurons resulted in a significant decrease in the density of thin dendritic spines, revealing the plausible biological mechanisms of *PACSI* in BD. However, further in vivo analyses (e.g. studies in transgenic mice) are needed to demonstrate the molecular mechanism of *PACSI* in the pathogenesis of BD. Fifth, our findings do not guarantee that the functional SNPs identified in this study are the most relevant SNPs for understanding susceptibility to bipolar disorder. This is a critical limitation of our study. However, considering that pinpointing functional (or causal) variants in the reported risk loci and elucidating their regulatory mechanisms remain challenging in the post-GWAS era, our findings may provide some new insights into the genetic mechanisms of bipolar disorder. Sixth, we characterized only three SNPs (rs10896081, rs3862386, and rs2027349) (Figs. 4, 5, and 6) in detail in our study. The major reasons that we characterized these three SNPs are as follows: (i) The ChIP-Seq data clearly showed that the corresponding TFs bound to genomic sequences containing these three SNPs in human brain tissues or neuronal cells; (ii) these three SNPs were characterized by strong DNase-Seq and histone modification signals; and (iii) the reporter gene assays of these three SNPs showed significant differences between different alleles, with the same effect direction. Characterization of more TF binding-disrupting SNPs will provide more insights into the genetic regulatory mechanisms of BD. Finally, our results suggested that two regulatory SNPs (rs10896081 and rs3862386) might act independently to regulate the potential target gene *PACSI*. However, more work is needed to demonstrate whether these two SNPs act independently or synergistically to regulate *PACSI*.

## Conclusions

In summary, we identified 16 functional SNPs in 9 reported BD risk loci and demonstrated the functional consequences of these SNPs. Our results revealed the

complex gene regulatory mechanisms of BD risk variants and provided potential targets for clinical drug development.

## Abbreviations

GWASs: Genome-wide association studies; BD: Bipolar disorder; ChIP-Seq: Chromatin immunoprecipitation sequencing; SNP: Single-nucleotide polymorphism; ASE: Allele-specific expression; eQTL: Expression quantitative trait loci; TF: Transcription factor; LD: Linkage disequilibrium; PWM: Position weight matrix; GWS: Genome-wide significant; FIMO: Find individual motif occurrences; HEK293T: Human embryonic kidney cell line; SH-SY5Y: Human neuroblastoma cell line; U251: Human glioblastoma cell line; CMC: Common mind consortium; GTEx: Genotype-tissue expression; LIBD: Lieber Institute for Brain Development; xQTL: xQTL map of the human brains; shRNA: Short hairpin RNA; RT-qPCR: Real-time quantitative PCR; KO: Knockout; CNVs: Copy number variations.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-022-01057-3>.

**Additional file 1: Supplementary Tables S2-S5 and Supplementary Figures S1-S5.**

**Additional file 2: Supplementary Table S1, Supplementary Tables S6-S9.**

## Acknowledgements

One of the brain eQTL datasets used in this study was generated as part of the Common Mind Consortium supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd, and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, R01-MH-075916, P50M096891, P50MH084053S1, R37MH057881 and R37MH057881S1, HHSN271201300031C, AG02219, AG05138, and MH06692. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories, and the NIMH Human Brain Collection Core. CMC Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Keisuke Hirai, Hiroyoshi Toyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner, and Barbara Lipska (NIMH). The Genotype-Tissue Expression (GTEx) project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. We thank Dr Thomas Whittington (Department of Medical Epidemiology and Biostatistics, Karolinska Institute) for sharing the curated PWM data with us. We thank Miss. Qian Li for her technical assistance.

## Authors' contributions

X.J.L. and M.L. conceived, designed, and supervised the whole study. R.C. performed most of the experiments, including reporter gene assays, knockdown of TFs, CRISPR-Cas9-mediated genome editing, and RT-qPCR. Z.H.Y., H.C., and X.C. performed the dendritic spine assays. J.W.L. and Y.X.H. conducted the functional genomic analyses, including the processing of ChIP-Seq data, derivation of TF binding motifs, DNase-Seq, and histone modifications. J.W.L. performed the ASE and eQTL analyses. R.C., J.W.L., and X.J.L. drafted the manuscript. M.L. and Z.J.Z. contributed to this work in study design, data interpretation, manuscript writing, and revision. All authors read this manuscript carefully, provided critical comments, and approved the final manuscript. The author(s) read and approved the final manuscript.

## Funding

This study was equally supported by the National Nature Science Foundation of China (U2102205 and 31970561 to X.J.L., 82071534 and 81871067 to H.C.,

31900414 to Y.X.H., 81830040 and 82130042 to Z.J.Z.) and the Distinguished Young Scientists grant of Yunnan Province (202001AV070006). Additionally, it was supported by the Western Light Innovative Research Team of the Chinese Academy of Sciences, the Key Research Project of Yunnan Province (202101AS070055), and the Science and Technology Program of Guangdong (No. 2018B030334001).

#### Availability of data and materials

The PWM data from Whittington et al. are accessible at the NCBI GEO website under the accession number GSE70770 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70770>) [44]. The ChIP-Seq data and histone modification data were downloaded from the ENCODE project (<https://www.encodeproject.org/>) [45]. The PsychENCODE brain eQTL data were downloaded from the PsychENCODE website (<http://resource.psychencode.org/>) [52]. The CMC brain eQTL data are accessible at the NCBI GEO website under the accession number GSE30272 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30272>) [53]. The Brain xQTL data were downloaded from the xQTL website (<http://mostafavilab.stat.ubc.ca/xQTLServer/>) [54]. The GTEx brain eQTL data were downloaded from the GTEx website (<https://gtexportal.org/home/>) [55]. The LiBD2 brain eQTL data were downloaded from the eQTL website (<http://eqtl.brainseq.org/phase2/eqtl/>) [56]. The PWM data, ChIP-Seq, DNase-Seq, and histone modification data of the 16 TF binding disrupting SNPs are available in additional file 1, figure S5.

#### Declarations

##### Ethics approval and consent to participate

All experiments were approved by the Animal Ethics Committee of the Kunming Institute of Zoology (License number: SMKX-2021-01-001) and conformed to National Advisory Committee for Laboratory Animal Research guidelines.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China. <sup>2</sup>Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650204, China. <sup>3</sup>Department of Neurology, Affiliated Zhongda Hospital, Southeast University, Nanjing, Jiangsu 210096, China. <sup>4</sup>Key Laboratory of Developmental Genes and Human Disease of Ministry of Education, Southeast University, Nanjing, Jiangsu 210096, China. <sup>5</sup>Zhongda Hospital, School of Life Sciences and Technology, Advanced Institute for Life and Health, Southeast University, Nanjing, Jiangsu 210096, China.

Received: 15 July 2021 Accepted: 11 May 2022

Published online: 20 May 2022

#### References

- Carvalho AF, Firth J, Vieta E. Bipolar Disorder. *N Engl J Med*. 2020;383(1):58–66. <https://doi.org/10.1056/NEJMra1906193>.
- Merikangas KR, Akiskal HS, Angst J, Greenberg PE, Hirschfeld RM, Petukhova M, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. *Arch Gen Psychiatry*. 2007;64(5):543–52. <https://doi.org/10.1001/archpsyc.64.5.543>.
- Crump C, Sundquist K, Winkleby MA, Sundquist J. Comorbidities and mortality in bipolar disorder: a Swedish national cohort study. *JAMA Psychiatry*. 2013;70(9):931–9. <https://doi.org/10.1001/jamapsychiatry.2013.1394>.
- Correll CU, Solmi M, Veronese N, Bortolato B, Rosson S, Santonastaso P, et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. *World Psychiatry*. 2017;16(2):163–80. <https://doi.org/10.1002/wps.20420>.
- Miller JN, Black DW. Bipolar disorder and suicide: a review. *Curr Psychiatry Rep*. 2020;22(2):6. <https://doi.org/10.1007/s11920-020-1130-0>.
- Hosang GM, Korszun A, Jones L, Jones I, Gray JM, Gunasinghe CM, et al. Adverse life event reporting and worst illness episodes in unipolar and bipolar affective disorders: measuring environmental risk for genetic research. *Psychol Med*. 2010;40(11):1829–37. <https://doi.org/10.1017/S003329170999225X>.
- Anand A, Koller DL, Lawson WB, Gershon ES, Nurnberger Jr. Genetic and childhood trauma interaction effect on age of onset in bipolar disorder: An exploratory analysis. *J Affect Disord*. 2015;179:1–5. <https://doi.org/10.1016/j.jad.2015.02.029>.
- Alloy LB, Abramson LY, Urosevic S, Walshaw PD, Nusslock R, Neeren AM. The psychosocial context of bipolar disorder: environmental, cognitive, and developmental risk factors. *Clin Psychol Rev*. 2005;25(8):1043–75. <https://doi.org/10.1016/j.cpr.2005.06.006>.
- Sylvia LG, Alloy LB, Hafner JA, Gauger MC, Verdon K, Abramson LY. Life events and social rhythms in bipolar spectrum disorders: a prospective study. *Behav Ther*. 2009;40(2):131–41. <https://doi.org/10.1016/j.jad.2012.01.038>.
- McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry*. 2003;60(5):497–502. <https://doi.org/10.1001/archpsyc.60.5.497>.
- Mendlewicz J, Rainer JD. Adoption study supporting genetic transmission in manic–depressive illness. *Nature*. 1977;268(5618):327–9. <https://doi.org/10.1038/268327a0>.
- Bertelsen A, Harvald B, Hauge M. A Danish twin study of manic–depressive disorders. *Br J Psychiatry*. 1977;130:330–51. <https://doi.org/10.1192/bjp.130.4.330>.
- Cichon S, Muhleisen TW, Degenhardt FA, Mattheisen M, Miro X, Strohmaier J, et al. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet*. 2011;88(3):372–81. <https://doi.org/10.1016/j.ajhg.2011.01.017>.
- Li HJ, Zhang C, Hui L, Zhou DS, Li Y, Zhang CY, et al. Novel risk loci associated with genetic risk for bipolar disorder among Han Chinese individuals: a genome-wide association study and meta-analysis. *JAMA Psychiatry*. 2021;78(3):320–30. <https://doi.org/10.1001/jamapsychiatry.2020.3738>.
- Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet*. 2019;51(5):793–803. <https://doi.org/10.1038/s41588-019-0397-8>.
- Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, 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>.
- Zhang C, Xiao X, Li T, Li M. Translational genomics and beyond in bipolar disorder. *Mol Psychiatry*. 2021;26(1):186–202. <https://doi.org/10.1038/s41380-020-0782-9>.
- Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74. <https://doi.org/10.1038/nature15393>.
- Shriner D, Adeyemo A, Gerry NP, Herbert A, Chen G, Doumatey A, et al. Transferability and fine-mapping of genome-wide associated loci for adult height across human populations. *PLoS One*. 2009;4(12):e8398. <https://doi.org/10.1371/journal.pone.0008398>.
- Chen G, Ramos E, Adeyemo A, Shriner D, Zhou J, Doumatey AP, et al. UGT1A1 is a major locus influencing bilirubin levels in African Americans. *Eur J Hum Genet*. 2012;20(4):463–8. <https://doi.org/10.1038/ejhg.2011.206>.
- Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*. 2002;3(4):299–309. <https://doi.org/10.1038/nrg777>.
- Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS. A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet*. 2015;47(8):955–61. <https://doi.org/10.1038/ng.3331>.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–7. <https://doi.org/10.1038/nature13645>.

24. Liu J, Tang W, Budhu A, Forgues M, Hernandez MO, Candia J, et al. A viral exposure signature defines early onset of hepatocellular carcinoma. *Cell*. 2020;182(2):317–28. <https://doi.org/10.1016/j.cell.2020.05.038>.
25. Ye J, Tucker NR, Weng LC, Clauss S, Lubitz SA, Ellinor PT. A functional variant associated with atrial fibrillation regulates PITX2c expression through TFAP2a. *Am J Hum Genet*. 2016;99(6):1281–91. <https://doi.org/10.1016/j.ajhg.2016.10.001>.
26. Zhao B, Li T, Yang Y, Wang X, Luo T, Shan Y, et al. Common genetic variation influencing human white matter microstructure. *Science*. 2021;372(6548):eabf3736. <https://doi.org/10.1126/science.abf3736>.
27. Cai X, Dong J, Lu T, Zhi L, He X. Common variants in MAEA gene contributed the susceptibility to osteoporosis in Han Chinese postmenopausal women. *J Orthop Surg Res*. 2021;16(1):38. <https://doi.org/10.1186/s13018-020-02140-4>.
28. Verma SS, Cooke Bailey JN, Lucas A, Bradford Y, Linneman JG, Hauser MA, et al. Epistatic gene-based interaction analyses for glaucoma in eMERGE and NEIGHBOR Consortium. *PLoS Genet*. 2016;12(9):e1006186. <https://doi.org/10.1371/journal.pgen.1006186>.
29. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. <https://doi.org/10.1038/s41467-017-01261-5>.
30. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun*. 2017;8(1):599. <https://doi.org/10.1038/s41467-017-00257-5>.
31. Griesemer D, Xue JR, Reilly SK, Ulirsch JC, Kukreja K, Davis JR, et al. Genome-wide functional screen of 3'UTR variants uncovers causal variants for human disease and evolution. *Cell*. 2021;184(20):5247–60. <https://doi.org/10.1016/j.cell.2021.08.025>.
32. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet*. 2014;46(3):234–44. <https://doi.org/10.1038/ng.2897>.
33. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*. 2019;179(7):1469–82. <https://doi.org/10.1016/j.cell.2019.11.020>.
34. Smillie CS, Biton M, Ordovas-Montanes J, Sullivan KM, Burgin G, Graham DB, et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell*. 2019;178(3):714–30. <https://doi.org/10.1016/j.cell.2019.06.029>.
35. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*. 2016;167(5):1369–84. <https://doi.org/10.1016/j.cell.2016.09.037>.
36. Neville MDC, Choi J, Lieberman J, Duan QL. Identification of deleterious and regulatory genomic variations in known asthma loci. *Respir Res*. 2018;19(1):248. <https://doi.org/10.1186/s12931-018-0953-2>.
37. Piao X, Yahagi N, Takeuchi Y, Aita Y, Murayama Y, Sawada Y, et al. A candidate functional SNP rs7074440 in TCF7L2 alters gene expression through C-FOS in hepatocytes. *FEBS Lett*. 2018;592(3):422–33. <https://doi.org/10.1002/1873-3468.12975>.
38. Ali MW, Patro CPK, Devall M, Dampier CH, Plummer SJ, Kuscus C, et al. A functional variant on 9p21.3 related to glioma risk affects enhancer activity and modulates expression of CDKN2B-AS1. *Hum Mutat*. 2021;42(10):1208–14. <https://doi.org/10.1002/humu.24244>.
39. Buckley MA, Woods NT, Tyrer JP, Mendoza-Fandino G, Lawrenson K, Hazelett DJ, et al. Functional analysis and fine mapping of the 9p22.2 ovarian cancer susceptibility locus. *Cancer Res*. 2019;79(3):467–81. <https://doi.org/10.1158/0008-5472.CAN-17-3864>.
40. Li Y, Ma C, Li W, Yang Y, Li X, Liu J, et al. A missense variant in NDUFA6 confers schizophrenia risk by affecting YY1 binding and NAGA expression. *Mol Psychiatry*. 2021;26(11):6896–911. <https://doi.org/10.1038/s41380-021-01125-x>.
41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75. Available from: <http://zzz.bwh.harvard.edu/plink/>.
42. Huo Y, Li S, Liu J, Li X, Luo XJ. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nat Commun*. 2019;10(1):670. <https://doi.org/10.1038/s41467-019-08666-4>.
43. Li S, Li Y, Li X, Liu J, Huo Y, Wang J, et al. Regulatory mechanisms of major depressive disorder risk variants. *Mol Psychiatry*. 2020;25(9):1926–45. <https://doi.org/10.1038/s41380-020-0715-7>.
44. Whittington T, Gao P, Song W, Ross-Adams H, Lamb AD, Yang Y, et al. Gene regulatory mechanisms underpinning prostate cancer susceptibility. *Nat Genet*. 2016;48(4):387–97. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70770>.
45. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74. Available from: <https://www.encodeproject.org/>.
46. Kong Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics*. 2011;98(2):152–3. Available from: <http://graphics.med.yale.edu/trim/>.
47. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25. Available from: <http://bowtie-bio.sourceforge.net/index.shtml>.
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. Available from: <http://samtools.sourceforge.net>.
49. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137. Available from: <http://liulab.dfci.harvard.edu/MACS/>.
50. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*. 1994;2:28–36. Available from: <https://meme-suite.org/meme/tools/meme>.
51. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27(7):1017–8. Available from: <https://meme-suite.org/meme/tools/fimo>.
52. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420):eaat8464. Available from: <http://resource.psych.encode.org/>.
53. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neurosci*. 2016;19(11):1442–53. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30272>.
54. Ng B, White CC, Klein HU, Sieberts SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nature Neurosci*. 2017;20(10):1418–26. Available from: <http://mostafavilab.stat.ubc.ca/xQTLServer/>.
55. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318–30. Available from: <https://gtexportal.org/home/>.
56. Collado-Torres L, Burke EE, Peterson A, Shin J, Straub RE, Rajpurohit A, et al. Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia. *Neuron*. 2019;103(2):203–16. Available from: <http://eqtl.brainseq.org/phase2/eqtl/>.
57. Castel SE, Aguet F, Mohammadi P, Consortium GT, Ardlie KG, Lapalainen T. A vast resource of allelic expression data spanning human tissues. *Genome Biol*. 2020;21(1):234. <https://doi.org/10.1186/s13059-020-02122-z>.
58. Thermo FS. shRNA design. RNAi Designer. 2022. <http://rnaidesigner.thermo-fisher.com/rnaiexpress/setOption.do?designOption=shrna&pid=3105315568901923019>.
59. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods*. 2001;25(4):402–8. <https://doi.org/10.1006/meth.2001.1262>.
60. Srivastava DP, Woolfrey KM, Penzes P. Analysis of dendritic spine morphology in cultured CNS neurons. *J Vis Exp*. 2011;53:e2794. <https://doi.org/10.3791/2794>.
61. Yang Z, Zhou D, Li H, Cai X, Liu W, Wang L, et al. The genome-wide risk alleles for psychiatric disorders at 3p21.1 show convergent effects on mRNA expression, cognitive function, and mushroom dendritic spine. *Mol Psychiatry*. 2020;25(1):48–66. <https://doi.org/10.1038/s41380-019-0592-0>.
62. Cai X, Yang ZH, Li HJ, Xiao X, Li M, Chang H. A human-specific schizophrenia risk tandem repeat affects alternative splicing of a human-unique

- isoform AS3MTd2d3 and mushroom dendritic spine density. *Schizophrenia Bull.* 2021;47(1):219–27. <https://doi.org/10.1093/schbul/sbaa098>.
63. Rodriguez A, Ehlenberger DB, Dickstein DL, Hof PR, Wearne SL. Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images. *PLoS One.* 2008;3(4):e1997. <https://doi.org/10.1371/journal.pone.0001997>.
  64. Wearne SL, Rodriguez A, Ehlenberger DB, Rocher AB, Henderson SC, Hof PR. New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience.* 2005;136(3):661–80. <https://doi.org/10.1016/j.neuroscience.2005.05.053>.
  65. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164. <https://doi.org/10.1093/nar/gkq603>.
  66. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature.* 2011;470(7333):279–83. <https://doi.org/10.1038/nature09692>.
  67. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA.* 2010;107(50):21931–6. <https://doi.org/10.1073/pnas.1016071107>.
  68. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science.* 2018;362(6420):eaat8127. <https://doi.org/10.1126/science.aat8127>.
  69. Konopaske GT, Lange N, Coyle JT, Benes FM. Prefrontal cortical dendritic spine pathology in schizophrenia and bipolar disorder. *JAMA Psychiatry.* 2014;71(12):1323–31. <https://doi.org/10.1001/jamapsychiatry.2014.1582>.
  70. Harrison PJ, Geddes JR, Tunbridge EM. The emerging neurobiology of bipolar disorder. *Trends Neurosci.* 2018;41(1):18–30. <https://doi.org/10.1176/appi.focus.17309>.
  71. Lee Y, Zhang Y, Kim S, Han K. Excitatory and inhibitory synaptic dysfunction in mania: an emerging hypothesis from animal model studies. *Exp Mol Med.* 2018;50(4):1–11. <https://doi.org/10.1038/s12276-018-0028-y>.
  72. Haggarty SJ, Karmacharya R, Perlis RH. Advances toward precision medicine for bipolar disorder: mechanisms & molecules. *Mol Psychiatry.* 2021;26(1):168–85. <https://doi.org/10.1038/s41380-020-0831-4>.
  73. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry.* 2008;13(2):197–207. <https://doi.org/10.1038/sj.mp.4002012>.
  74. Psychiatric GCBWDG. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet.* 2011;43(10):977–83. <https://doi.org/10.1038/ng.943>.
  75. Chen DT, Jiang X, Akula N, Shugart YY, Wendland JR, Steele CJ, et al. Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. *Mol Psychiatry.* 2013;18(2):195–205. <https://doi.org/10.1038/mp.2011.157>.
  76. Green EK, Hamshere M, Forty L, Gordon-Smith K, Fraser C, Russell E, et al. Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Mol Psychiatry.* 2013;18(12):1302–7. <https://doi.org/10.1038/mp.2012.142>.
  77. Muhleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J, et al. Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat Commun.* 2014;5:3339. <https://doi.org/10.1038/ncomms4339>.
  78. Ikeda M, Takahashi A, Kamatani Y, Okahisa Y, Kunugi H, Mori N, et al. A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol Psychiatry.* 2018;23(3):639–47. <https://doi.org/10.1038/mp.2016.259>.
  79. Hou L, Bergen SE, Akula N, Song J, Hultman CM, Landen M, et al. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. *Hum Mol Genet.* 2016;25(15):3383–94. <https://doi.org/10.1093/hmg/ddw181>.
  80. Liu F, Wu D, Wang X. Roles of CTCF in conformation and functions of chromosome. *Semin Cell Dev Biol.* 2019;90:168–73. <https://doi.org/10.1016/j.semcdb.2018.07.021>.
  81. Wan L, Molloy SS, Thomas L, Liu G, Xiang Y, Rybak SL, et al. PACS-1 defines a novel gene family of cytosolic sorting proteins required for trans-Golgi network localization. *Cell.* 1998;94(2):205–16. [https://doi.org/10.1016/S0092-8674\(00\)81420-8](https://doi.org/10.1016/S0092-8674(00)81420-8).
  82. Dikeakos JD, Thomas L, Kwon G, Elferich J, Shinde U, Thomas G. An inter-domain binding site on HIV-1 Nef interacts with PACS-1 and PACS-2 on endosomes to down-regulate MHC-I. *Mol Biol Cell.* 2012;23(11):2184–97. <https://doi.org/10.1091/mbc.E11-11-0928>.
  83. Burgert T, Schmidt V, Caglayan S, Lin F, Fuchtbauer A, Fuchtbauer EM, et al. SORLA-dependent and -independent functions for PACS1 in control of amyloidogenic processes. *Mol Cell Biol.* 2013;33(21):4308–20. <https://doi.org/10.1128/MCB.00628-13>.
  84. Schuurs-Hoeijmakers JH, Oh EC, Vissers LE, Swinkels ME, Gilissen C, Willemssen MA, et al. Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. *Am J Hum Genet.* 2012;91(6):1122–7. <https://doi.org/10.1016/j.ajhg.2012.10.013>.
  85. Zainabadi K, Benyamini P, Chakrabarti R, Veena MS, Chandrasekharappa SC, Gatti RA, et al. A 700-kb physical and transcription map of the cervical cancer tumor suppressor gene locus on chromosome 11q13. *Genomics.* 2005;85(6):704–14. <https://doi.org/10.1016/j.ygeno.2005.02.014>.
  86. French JD, Ghossaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, et al. Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am J Hum Genet.* 2013;92(4):489–503. <https://doi.org/10.1016/j.ajhg.2013.01.002>.
  87. Chatterjee S, Karasaki KM, Fries LE, Kapoor A, Chakravarti A. A multi-enhancer RET regulatory code is disrupted in Hirschsprung disease. *Genome Res.* 2021;31:2199–208. <https://doi.org/10.1101/gr.275667.121>.
  88. Shidal C, Shu X, Wu J, Wang J, Huang S, Long J, et al. Functional genomic analyses of the 21q22.3 locus identifying functional variants and candidate gene YBEY for breast cancer risk. *Cancers.* 2021;13(9):2037. <https://doi.org/10.3390/cancers13092037>.
  89. Zhu DL, Chen XF, Hu WX, Dong SS, Lu BJ, Rong Y, et al. Multiple functional variants at 13q14 risk locus for osteoporosis regulate RANKL expression through long-range super-enhancer. *J Bone Miner Res.* 2018;33(7):1335–46. <https://doi.org/10.1002/jbmr.3419>.
  90. Fortini BK, Tring S, Plummer SJ, Edlund CK, Moreno V, Bresalier RS, et al. Multiple functional risk variants in a SMAD7 enhancer implicate a colorectal cancer risk haplotype. *Plos One.* 2014;9(11):e111914. <https://doi.org/10.1371/journal.pone.0111914>.
  91. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin YM, Albu M, et al. The human transcription factors. *Cell.* 2018;172(4):650–65. <https://doi.org/10.1016/j.cell.2018.01.029>.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

