Functional Genomics Identify a Regulatory Risk Variation rs4420550 in the 16p11.2 Schizophrenia-Associated Locus

Supplement 1

Supplemental Methods and Materials

Expression quantitative trait loci (eQTL) analysis in public brain datasets

Five gene expression datasets of genome-wide genotype data and polyA$^+$ RNA-Seq data of the dorsolateral prefrontal cortex (DLPFC) are included in the current expression quantitative trait loci (eQTL) analyses: BrainSeq Phase 1 (1), Brain xQTL (2), CommonMind (3), Genotype-Tissue Expression project (GTEx) (4), Psychiatric Encyclopedia of DNA Elements (PsychENCODE) (5).

The eQTL association results from 412 subjects were retrieved from BrainSeq Phase 1 (http://eqtl.brainseq.org/phase1/eqtl/), which were analyzed through linear regression adjusting for sex, ancestry, and expression heterogeneity (principal components) by the original authors.

From Brain xQTL (http://mostafavilab.stat.ubc.ca/xQTLServe/), eQTL association results between each SNP and mRNA levels of genes (estimated with Spearman's rank correlation) in 494 individuals were collected for the current study.

CommonMind (https://www.synapse.org/#!Synapse:syn2759792/wiki/69613) contains eQTL association results from 467 subjects, which were calculated according to the formula: adjusted gene expression ~ SNP + ancestry vectors + diagnosis.

The eQTL association results of frontal cortex tissues (BA9) from 175 individuals from GTEx (https://www.gtexportal.org/) were also used in the current analysis. These results were derived through linear regression analyses covaring top three genotyping principal components, gender, genotyping platforms, and additional covariates.

The eQTL summary data of 1,387 individuals in PsychENCODE project (6) were downloaded from the PsychENCODE Integrative Analysis website (http://resource.psychencode.org/) (7). The eQTL analysis was conducted by adjusting for diagnosis, study, age, age2, RIN, RIN2, sex, tissue, PMI, 20 ancestry PCs, and 100 hidden covariates (7). It should be noted that individuals in the PsychENCODE came from CommonMind, CommonMind-HBCC, BrainGVEX, LIBD and BipSeq *etc*.

SMR integrative analysis

Zhu *et al.* previously developed the summary data-based Mendelian randomization (SMR) analysis to prioritize risk genes, through integrating summary statistics from independent GWAS and eQTL datasets to assess the pleiotropic effects of SNPs on diagnosis and mRNA expression levels (8).

They further developed a multi-SNP-based SMR analysis that simultaneously considers multiple SNPs at a single cis-eQTL locus (9). For the current SMR integrative analysis, the eQTL *p*-value threshold was set at 1.00×10^{-5}, and default values were used for other parameters. The genes having $P_{\text{SMR-multi}}$ lower than 1.00×10^{-4} in multi-SNP-based SMR test and no evidence of significant heterogeneity ($P_{\text{HEIDI}} > 0.01$) were considered as schizophrenia susceptibility genes. In the current SNR analyses, we integrated the schizophrenia GWAS (40,675 cases and 64,643 controls) (10) with the DLPFC eQTL datasets from BrainSeq Phase 1 (1), Brain xQTL (2), CommonMind (3), GTEx frontal cortex (4), and PsychENCODE (5) respectively.

Cell culture

The HEK293T (human embryonic kidneys 293T) and SK-N-SH (human neuroblastoma) cell lines were originally from the Kunming Cell Bank, Kunming Institute of Zoology, Chinese Academy of Sciences. Both cells were cultured in a humidified 5% CO_2 incubator at 37 °C in Dulbecco's Modified Eagle's Medium (DMEM) basic (Gibco) supplemented with 10% fetal bovine serum (FBS) (Biological Industries), 1% non-essential amino acids, 1% Sodium pyruvate and 1% penicillin-streptomycin. PCR detection was regularly performed to ensure there were no mycoplasma contamination.

Luciferase plasmid construction and in vitro luciferase reporter assay

Based on the DNase-Seq and histone modification data in cells and brain tissues, a 589-bp rs4420550-centered region was selected for enhancer activity assays. The DNA fragment containing rs4420550 was amplified from genomic DNA of HEK293T cells using the primers 5'-CTGGAGGTGGAGGTTGTG-3' (forward) and 5'-AGGTCCCTGGGAGATGCT-3' (reverse). Site-directed mutagenesis was employed to generate sequences containing either A or G allele at rs4420550. A 477-bp DNA fragment spanning rs12691307 (a noncoding variation associated with schizophrenia but exhibits few functional annotations in our prediction analyses) was also cloned for comparison (primers, 5'-TGGAGGAGCAACAAAGATTCAG-3' (forward) and 5'-ATGGTGGCTCACGCCTGTA-3' (reverse)).

The DNA fragments were cloned into the pGL3-promoter vector upstream of an SV40 promoter, and the recombinant clones were verified through bi-directional DNA sequencing to make sure no *de novo* mutation was introduced. All plasmids were accurately quantified. Equal amounts of each

plasmid were transiently transfected into HEK293T and SK-N-SH cells together with pRL-TK plasmid (a standard reporter for internal control) using Lipofectamine 2000 (Thermo Scientific, USA) according to the manufacturer's instructions. Twenty-four hours post-transfection, cells were lysed and luciferase activity was determined using the Dual-Luciferase Reporter Assay System on a GloMax 96 Luminometer (Promega, USA). The firefly luciferase activity was normalized to that of Renilla luciferase to control for differences in transfection efficiency. All assays were performed in at least three independent experiments, and each experimental condition was done in triplicates. Two-tailed t-tests were performed for statistical analyses.

Chromosome conformation capture (3C) analysis

The chromosome conformation capture (3C) assay was performed following previous studies (11,12). HEK293T cells were cultured and harvested in log phase growth. A total of 1×10^7 cells in 10 mL of DMEM culture medium were fixed using 1% formaldehyde at room temperature for 10 min. The crosslinking was then stopped by glycine, and the fixed cells were lysed in 5 mL lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA-630 with Protease Inhibitor Cocktail from Thermo Scientific, USA) for 15 min at 4 ºC with gentle pipetting. The mixture was centrifuged, and the nuclei pellet was resuspended in 500 μL of 1.2× restriction enzyme buffer with 0.3% SDS and incubated at 37 ºC for 1 h with shaking at 900 rpm. Following this, the SDS was sequestered through incubation with 2% Triton X-100 at 37 ºC for another hour with shaking, and pellets were digested with 400 U restriction enzyme *EcoR*I (NEB, USA) for at least 16 h. NEB digestion was then stopped by adding SDS to 1.6% and incubating at 65 ºC for 30 min. The digested chromatin was diluted in 6.125 mL of 1.15 × ligation buffer (NEB, USA), and residual SDS was sequestered by adding Triton X-100 to 2% and incubating at 37 ºC for 1 h with shaking. Subsequently, 100 U high concentration T4 DNA ligase (Promega, USA) was added to the diluted chromatin, and incubated for 4 h at 16 ºC followed by 30 min at room temperature. Proteinase K (500 μg) was then added to the chromatin mixture at 65 ºC overnight to reverse crosslinks, and RNA was removed by RNase A (300 μg) treatment for 60 min at 37 ºC. The 3C library was purified through phenol-chloroform extraction and ethanol precipitation. TaqMan RT-qPCR was performed to determine the interaction frequency between target sites of interest (specific primers listed in **Supplemental Table S2**). TaqMan RT-qPCR was performed using the following cycling conditions: 95 ºC for 10 min; and 45 cycles of 15 sec at 95 ºC and 60 sec at 60 ºC. Sanger sequencing was performed on the purified PCR products to determine the accuracy

of each chimeric DNA. To normalize primer efficiency, control PCR templates were generated by digestion and random ligation of bacterial artificial chromosomes covering 16p11.2 region (CTC-632p24 and RP11-114A14 from Thermo Scientific). All the 3C primers and internal control primer were tested for amplification efficiencies in the BAC libraries, and a standard curve was made for each primer (**Supplemental Table 11**). For the calculation of the interaction frequencies, the value for each detectable locus was calculated using the parameters of the standard curve as follows: value=$10^{(Ct-\text{intercept})/\text{slope}}$, and the values were then normalized to the value of internal control to generate relative interaction frequency.

Precise genome editing by CRISPR/Cas9

Homology-directed repair (HDR)-mediated genome editing by CRISPR/Cas9 was performed following a previous study with modifications (13). Plasmid pL-CRISPR.EFS.GFP was purchased from Addgene (#57818). To minimize potential off-target effect of CRISPR/cas9, eSpCas9 (1.1) was used instead of the wild-type cas9. The open reading frame of eSpCas9 (1.1) including 3×FLAG tag was amplified from plasmid eSpCas9 (1.1) (Addgene, #71814) and sub-cloned into pL-CRISPR.EFS.GFP to replace the wild-type cas9 with BamHI and NheI restriction sites (this construct was named pL-CRISPR-E.EFS.GFP). Protospacer sequence of CRISPR/Cas9 targeting rs4420550 (5'-GATGTTGCTAGGAGCTGACC-3') was designed using Optimized CRISPR Design (http://crispr.mit.edu), and the selected sequence contained at least 3 mismatches when aligned with all potential off-target sites. Annealed oligomers were cloned into the pL-CRISPR-E.EFS.GFP plasmid. To precisely edit the single base pair at rs4420550 locus, a single-stranded oligodeoxynucleotide (ssODN) was synthesized as a homology directed repair template, and the ssODN sequence was 5'-gttttttgtgaaattattttgctcaacaggacaccaccatggccgtgctgttggtgccaggCcagctcctagcaacatcaaggcttctggagtgaggg tgcaaaccagctcccagctggcagc-3'.

HEK293T cells were seeded into 100 mm plates one day before transfection. The pL-CRISPR-E.EFS.GFP construct and the ssODN were transiently co-transfected using Lipofectamine 3000 (Thermo Scientific, USA) according to the manufacturer's instructions. Forty-eight hours after transfection, cells positive for GFP fluorescence were sorted with a flow cytometer, and cultured for 5-7 days to form distinct colonies derived from single cells. Genomic DNA was then extracted from each colony for amplifying the target region containing rs4420550, and Sanger sequencing was performed to confirm the genomic editing. Overall, eight single cell colonies were selected after flow

cytometry for subsequent analysis. Four colonies carried the [A/A] genotype at rs4420550, and the other four carried the [G/G] genotype. These colonies were confirmed to have the same genomic sequence only except for the rs4420550 locus, and no additional mutation was introduced besides rs4420550 during the genome editing.

CRISPR/Cas9 off-target assay

Off-target effects of the CRISPR/cas9 editing was analyzed as previously described (14). Twelve potential off-target sites showing high homology with the rs4420550 locus were validated. T7EN1 assay was performed to examine the cleavage of potential off-target sites. Specifically, a total of 200 ng purified PCR products were denatured and reannealed in 1×T7 Endonuclease I Reaction Buffer (Vazyme) in 20 μL volume using a thermocycler with the following program: 95 °C, 5 min; 95-85 °C at −2 °C/sec; 85-25 °C at −0.1 °C/sec; hold at 4 °C. 1 μL of T7 Endonuclease I (Vazyme) were then added to the PCR products and incubated at 37 °C for 20 min for hybridization. The PCR products digested by T7EN1 were separated using a 2% agarose gel and images were captured by Tanon 5200 Multi.

Real-time quantitative PCR

Total RNAs were extracted from cultured cells using the TRIzol Reagent. An aliquot of 2 μg total RNAs were synthesized to cDNA using the RevertAid First Strand cDNA Synthesis Kit in a 20 μL reaction mixture containing Random Hexamer Primer, RevertAid M-MuLV RT, RiboLock RNase Inhibitor, 5×Reaction Buffer and 10 mM dNTP Mix (Thermo Scientific, USA). The mRNA expression was quantified through real-time quantitative PCR (RT-qPCR) using the ABI PRISM 7900 real-time PCR system (Applied Biosystems) as described (15-17). *RPS13* was used as the reference gene to normalize the amplification signal between different wells and amount of input cDNA. The primer sequences used for amplifying *RPS13* and genes of interest in the 16p11.2 region were shown in **Supplemental Table S1**. The data of relative mRNA expression were presented as the means of $2^{-\Delta\Delta Ct}$. Statistical tests between different groups were conducted using two-tailed t-test.

RNA-Seq analysis in HEK293T cells

Fastq files of 150-bp paired-end reads of cDNA sequences were obtained using the Illumina NovaSeq 6000. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to

assess the quality of the sequence reads, and those with poor quality were filtered using trimmomatic-0.36 (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:50 HEADCROP:15) (18). The clean reads were then aligned to the human genome reference (GRCh38) using HISAT2 (19). Gene-level read counts were calculated through htseq-count based on the gene annotation gtf file GRCh38.91 (20). The Wald test in DESeq2 (21) was used to analyze the expression differences between experimental groups. Genes whose expression levels were significantly altered with a false discovery rate (FDR)<0.05 were identified.

Western blotting

Anti-p44/42 MAPK (Erk1/2) antibody (#9102) was purchased from Cell Signaling Technology; anti-TAOK2 (#21188-1-AP), anti-HIRIP3 (#14992-1-AP) and anti-ALDOA antibodies (#11217-1-AP) were purchased from Proteintech Group; anti-INO80E antibody (#PA5-60259) was purchased from Thermo Scientific; anti-KCTD13 antibody (#HPA043524) was purchased from Sigma-Aldrich; anti-α-tubulin antibody (#sc-8035) was purchased from Santa Cruz. The protocol for western blotting was described previously (22). In brief, HEK293T cells were washed with cold PBS and lysed using RIPA lysis buffer supplemented with protease inhibitors cocktail (Thermo Scientific, USA). Total proteins were then isolated and concentrations were determined using a BCA protein assay kit (Thermo Scientific, USA). The denatured protein samples were loaded, separated on an 8% SDS-PAGE gel and wet transferred to polyvinylidene fluoride (PVDF) membranes (EMD Millipore, USA). The transblots were blocked with 5% non-fat milk and then incubated with primary antibodies at 4 °C overnight. After incubation, the transblots were rinsed 3 times (10 min each time) with Tris-buffered saline solution containing 0.1% Tween 20, and then incubated with peroxidase-linked secondary antibodies for 60 min at room temperature. Finally, the transblots were incubated with SuperSignal West FemtoChemiluminescent Substrate (Thermo Scientific, USA) for detection of HRP in a dark chamber, and images were captured using Tanon 5200 imaging system. The density of the bands was quantified by Image-J software. The references for the dilution of all antibodies are 1:1000.

Assay for transposase-accessible chromatin followed by sequencing (ATAC-Seq)

The ATAC-Seq libraries of HEK293T cells were prepared by using the TruePrep DNA Library Prep Kit (Vazyme Biotech, China) according to the manufacturer's instruction with modifications. In brief, 3×10^4 HEK293T cells were lysed in lysis buffer [10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM

$MgCl_2$ and 0.1% (v/v) Igepal CA-630] on ice for 10 min to release the nuclei. The mixture was centrifuged at 4 ºC at 500× g for 5 min and the supernatant was discarded as much as possible. The nuclei pellet was then incubated with a reaction mix containing Tn5 transposome at 37 ºC for 30 min. DNA was purified immediately after transposition by using a Qiagen MinElute PCR Purification Kit. The transposed DNA fragments were PCR-amplified using the following conditions: 72 °C for 3 min, 98 °C for 30 sec, and thermocycling for 15 cycles at 98 °C for 15 sec, 60 °C for 30 sec and 72 °C for 30 sec, followed by 72 °C for 5 min. The PCR products were immediately purified with 1.2×AMPure (Beckman) beads to obtain ATAC-Seq libraries. The libraries quality was assessed by Agilent 2100 Bioanalyzer before sequencing.

The raw 150-bp paired-end reads of the ATAC-Seq libraries from HiSeq X Ten sequencer underwent similar quality control processes as those for RNA-seq results using FastQC and trimmomatic-0.36 (18). Bowtie2 (-X 2000 --sensitive-local --dovetail) was used to align trimmed reads to the GRCh37.p13 build of the human genome. Potential PCR duplicate reads were marked using MarkDuplicates program from the Picard tool (https://broadinstitute.github.io/picard/). Reads from mitochondrial DNA and the Y chromosome were filtered using SAMtools (http://www.htslib.org/). Eventually, reads mapped in pair for the further calling peaks were proceeded for further analyses. Peaks were called in the each sample using MACS2 (23) with parameters: --nomodel --shift -100 --extsize 200. Narrow peaks mapped to the blacklisted regions (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/) were excluded using bedtools (https://bedtools.readthedocs.io/en/latest/). Peaks called from different samples were then merged using the mergePeaks program (-d 2000) in the HOMER tool (http://homer.ucsd.edu/homer/introduction/install.html). The reads counts of each peak were quantified using featureCounts based on the chromosome position of merged peaks (24). The Wald test in DESeq2 was used to identify differently accessible peaks between groups (21). To visualize peaks, bins per million mapped reads (BPM) was used to quantify the coverage of peak using bamCoverage (binSize=10) in the deeptools (https://deeptools.readthedocs.io/en/develop/). The ATAC-Seq tag intensity around transcriptional start site (TSS) (±3 kb window) was calculated and plotted using computeMatrix and plotHeatmap programs in the deeptools.

Electrophoretic mobility shift assay

41-base pair oligonucleotide probe containing A or G allele at rs4420550, with 3'-end labeled with biotin, was purchased from Generay Biotechnology (The nucleotide sequences of the double-

stranded oligonucleotides with either A or G allele were shown in **Supplemental Table S4**). Nuclear extracts from HEK293T cells were isolated using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific, USA). Electrophoretic mobility shift assay was performed using a chemiluminescent EMSA kit (Beyotime, China) according to manufacturer's protocol. Briefly, 15 μg nuclear extracts and 50 fM biotin-labeled oligonucleotides were incubated at room temperature for 20 min for the binding reaction mix. The binding mixtures were then resolved in 6 % nondenaturing polyacrylamide gel and transferred onto a nylon membrane. The membrane was then cross-linked using ultraviolet, blocked and incubated with avidin-HRP enzyme, and incubated with chemiluminescence substrates. Images were captured by Tanon 5200 Multi.

Figure S1. The schizophrenia risk SNPs at 16p11.2 constitute two LD blocks. The LD map is generated using genotype data of European samples from the 1000 Genomes Project by Haploview v4.1 program (25). The SNPs rs12691307 and rs4420550 are located in LD block 1.

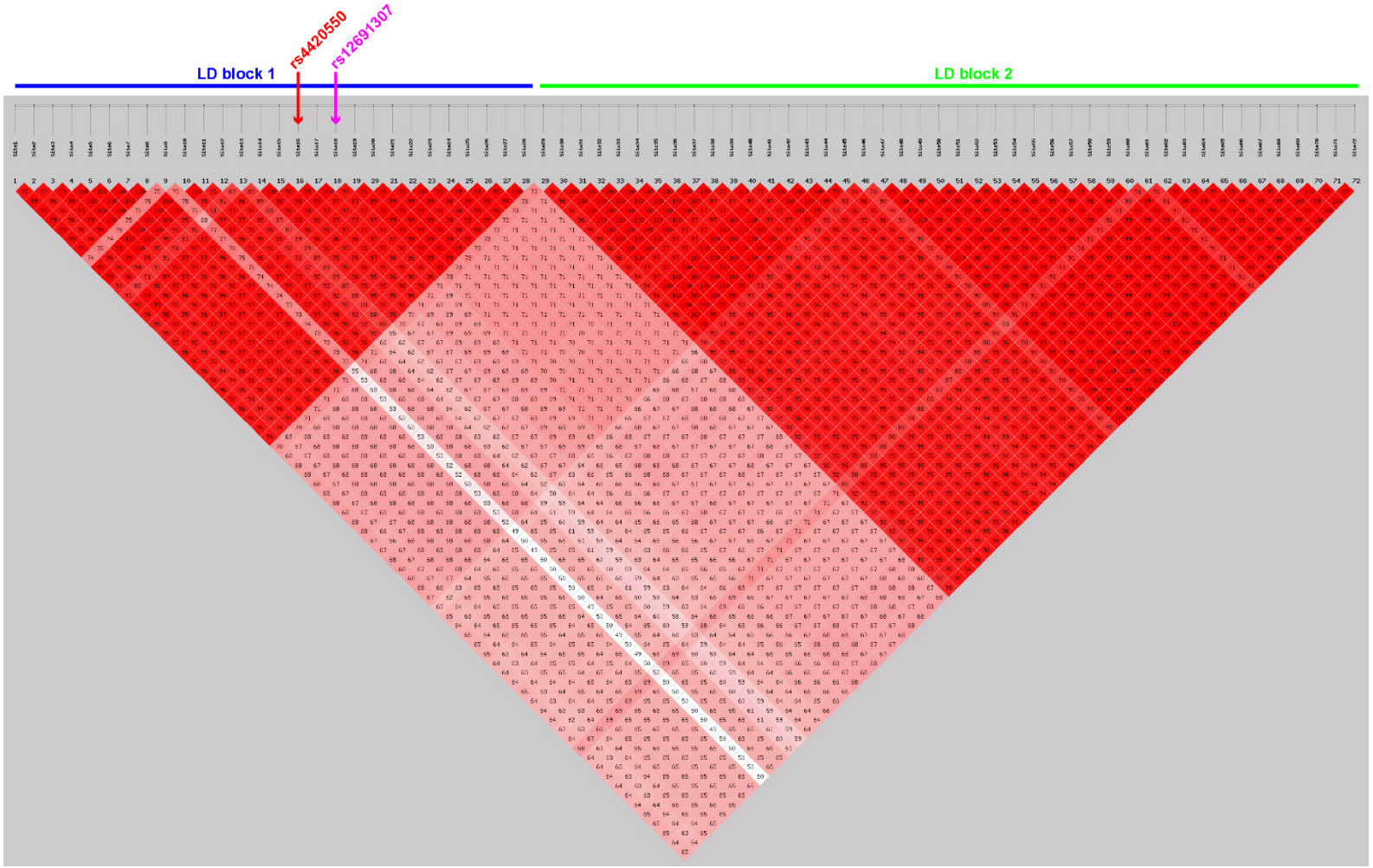


Figure S2. rs4420550 is predicted to be located in the DNA enhancer region in RegulomeDB (https://www.regulomedb.org/regulome-search/) (26).

Histone modifications — Filter: brain

Method	Location	Chromatin State	Tissue Group	Tissue	Reference
ChromHMM	chr16:29938200..29939200	Flanking Active TSS	Brain	Fetal Brain Female	REMC
ChromHMM	chr16:29938600..29939000	Enhancers	Brain	Brain Substantia Nigra	REMC
ChromHMM	chr16:29938600..29939000	Enhancers	Brain	Brain Hippocampus Middle	REMC
ChromHMM	chr16:29938600..29939400	Flanking Active TSS	Brain	Brain Germinal Matrix	REMC
ChromHMM	chr16:29938600..29939600	Enhancers	Brain	Fetal Brain Male	REMC
ChromHMM	chr16:29938600..29939600	Enhancers	Brain	Brain Cingulate Gyrus	REMC
ChromHMM	chr16:29938600..29939600	Enhancers	Brain	Brain Angular Gyrus	REMC
ChromHMM	chr16:29938600..29939800	Enhancers	Brain	Brain Dorsolateral Prefrontal Cortex	REMC
ChromHMM	chr16:29938800..29939400	Enhancers	Brain	Brain Inferior Temporal Lobe	REMC
ChromHMM	chr16:29938800..29939400	Enhancers	Brain	Brain Anterior Caudate	REMC

Figure S3. Functional predictions of the five SNPs in RegulomeDB (beta version) (26). Rs4420550 is predicted to locate in enhancer regions in many human tissues (n=61); rs12716972 and rs12716973 are predicted to locate in the Active TSS regions in many human tissues (n=121 and 122, respectively); rs12691307 and rs4424923 are predicted to locate in the Quiescent/Low regions in many human tissues (n=113 and 96, respectively).

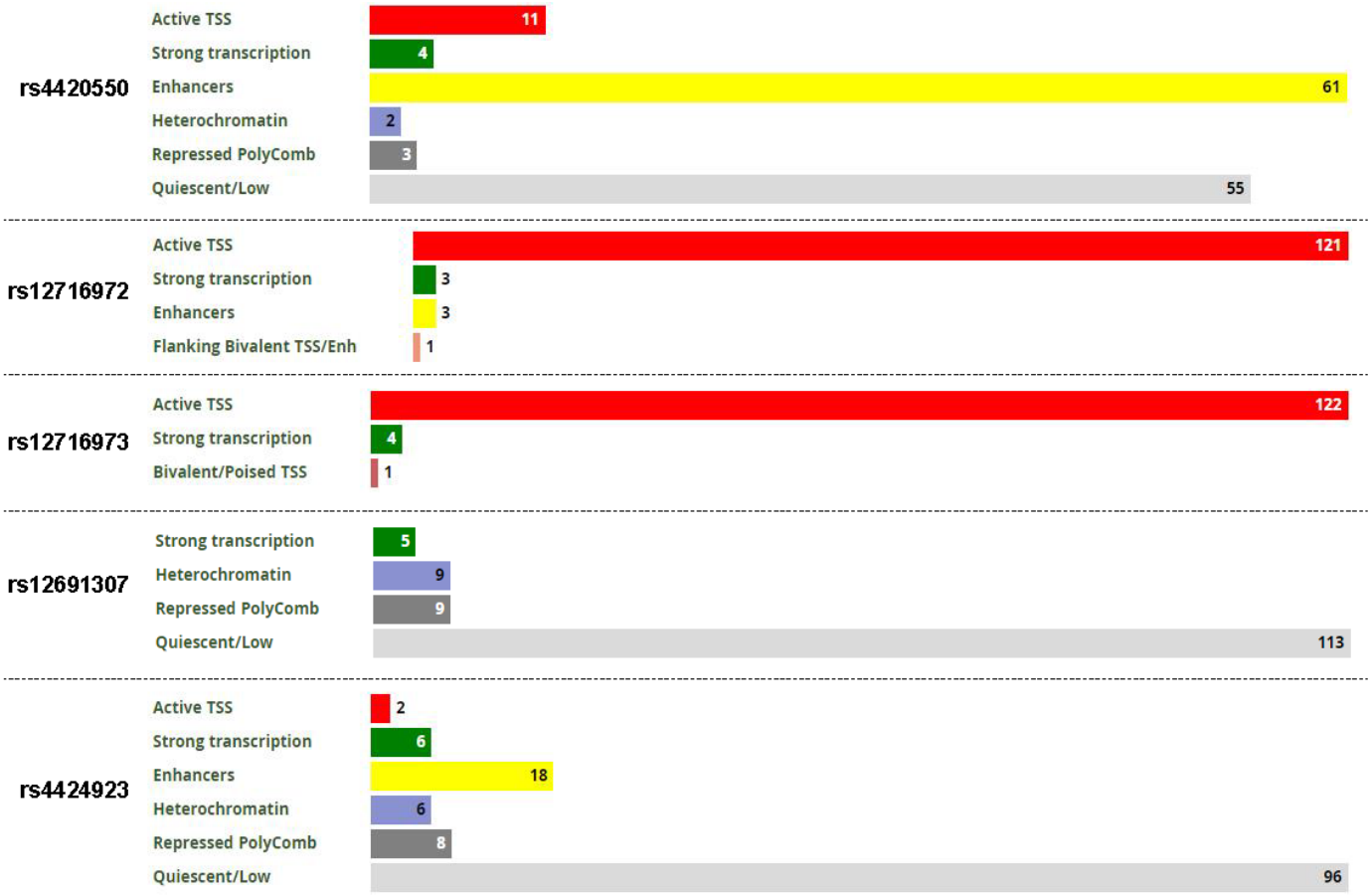


Figure S4. Functional predictions of the schizophrenia risk SNPs at 16p11.2 using HaploReg v4.1 dataset (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) (27).

chr	pos (hg38)	LD (r^2)	LD (D')	variant	Ref	Alt	AFR freq	AMR freq	ASN freq	EUR freq	SiPhy cons	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	NHGRI/EBI GWAS hits	GRASP QTL hits	Selected eQTL hits	GENCODE genes	dbSNP func annot
16	29913101	0	0	rs12444978	A	T	0.16	0.40	0.37	0.51						5 altered motifs			72 hits	ASPHD1	intronic
16	29913584	0	0	rs7201384	A	G	0.40	0.43	0.37	0.51				IPSC		ATF3			65 hits	ASPHD1	intronic
16	29914124	0	0	rs8059619	C	T	0.16	0.40	0.37	0.51						6 altered motifs			70 hits	ASPHD1	intronic
16	29916381	0	0	rs10083738	G	A	0.16	0.40	0.37	0.51								3 hits	71 hits	ASPHD1	intronic
16	29918407	0	0	rs8052502	C	T	0.16	0.40	0.37	0.51						5 altered motifs			72 hits	ASPHD1	intronic
16	29920578	0	0	rs4407079	G	A	0.16	0.40	0.37	0.51		BRN	BRN	BRN,BRN		4 altered motifs		2 hits	74 hits	CTD-2574D22.4	intronic
16	29920743	0	0	rs4609871	C	T	0.40	0.42	0.37	0.51		BRN	BRN	BRN,BRN		GR,YY1		8 hits	72 hits	CTD-2574D22.4	intronic
16	29921370	0	0	rs9936474	T	C	0.16	0.40	0.37	0.51		BRN	BRN	BRN		NRSF			72 hits	CTD-2574D22.4	intronic
16	29923433	0	0	rs12596042	T	A	0.16	0.40	0.37	0.50			6 tissues			Zfp105			72 hits	KCTD13	intronic
16	29923498	0	0	rs11150574	C	T	0.16	0.40	0.37	0.51			BLD, BRN						72 hits	KCTD13	intronic
16	29925333	0	0	rs12716972	A	G	0.41	0.42	0.37	0.51		23 tissues	GI, SPLN	13 tissues		SIX5,Smad3			64 hits	KCTD13	intronic
16	29926331	0	0	rs12716973	G	A	0.16	0.40	0.37	0.51		24 tissues		53 tissues	22 bound proteins	4 altered motifs			66 hits	KCTD13	
16	29927499	0	0	rs4420550	A	G	0.37	0.42	0.37	0.51		7 tissues	16 tissues	10 tissues		4 altered motifs			70 hits	CTD-2574D22.2	
16	29927972	0	0	rs4424923	T	C	0.41	0.43	0.37	0.51		SKIN, BRN	6 tissues	IPSC		NRSF,PLAG1,Zfx			69 hits	CTD-2574D22.2	
16	29928556	0	0	rs12691307	A	G	0.35	0.39	0.35	0.48						GCNF	1 hit		65 hits	CTD-2574D22.2	
16	29932046	0	0	rs12919683	T	C	0.17	0.39	0.36	0.50			BLD			NF-I			63 hits	3.1kb 3' of CTD-2574D22.2	
16	29932138	0	0	rs12933498	C	A	0.17	0.40	0.37	0.51			BLD			Dobox4			72 hits	3.2kb 3' of CTD-2574D22.2	
16	29933511	0	0	rs12716974	G	A	0.17	0.40	0.37	0.51			GI			11 altered motifs		5 hits	75 hits	4.6kb 3' of CTD-2574D22.2	
16	29935349	0	0	rs11649612	T	C	0.17	0.40	0.37	0.51						BDP1			75 hits	5.5kb 5' of CTD-2574D22.1	
16	29935387	0	0	rs11642933	G	A	0.16	0.40	0.37	0.51				BLD		DEC,Myc,Sin3Ak-20			73 hits	5.5kb 5' of CTD-2574D22.1	
16	29935624	0	0	rs4527039	T	C	0.38	0.42	0.37	0.50						Nrf-2,TCF11::MafG,ZID			68 hits	5.3kb 5' of CTD-2574D22.1	
16	29937080	0	0	rs4788190	G	A	0.39	0.43	0.37	0.51						Nkx2,Nkx3			72 hits	3.8kb 5' of CTD-2574D22.1	
16	29939508	0	0	rs12716975	C	T	0.16	0.40	0.37	0.51			LIV			NF-E2			72 hits	1.4kb 5' of CTD-2574D22.1	
16	29943333	0	0	rs4402589	T	G	0.41	0.41	0.37	0.50			14 tissues	BLD		8 altered motifs			73 hits	TMEM219	
16	29943634	0	0	rs3924855	A	G	0.40	0.42	0.37	0.51		5 tissues	16 tissues	20 tissues	14 bound proteins	CAC-binding-protein,ERalpha-a			74 hits	TMEM219	

Figure S5. Association of rs4420550 with MAPK3 mRNA expression in multiple human tissues in GTEx dataset (http://www.gtexportal.org/home/) (4). The effect size (NES) of eQTL association is calculated based on the A-allele of rs4420550.

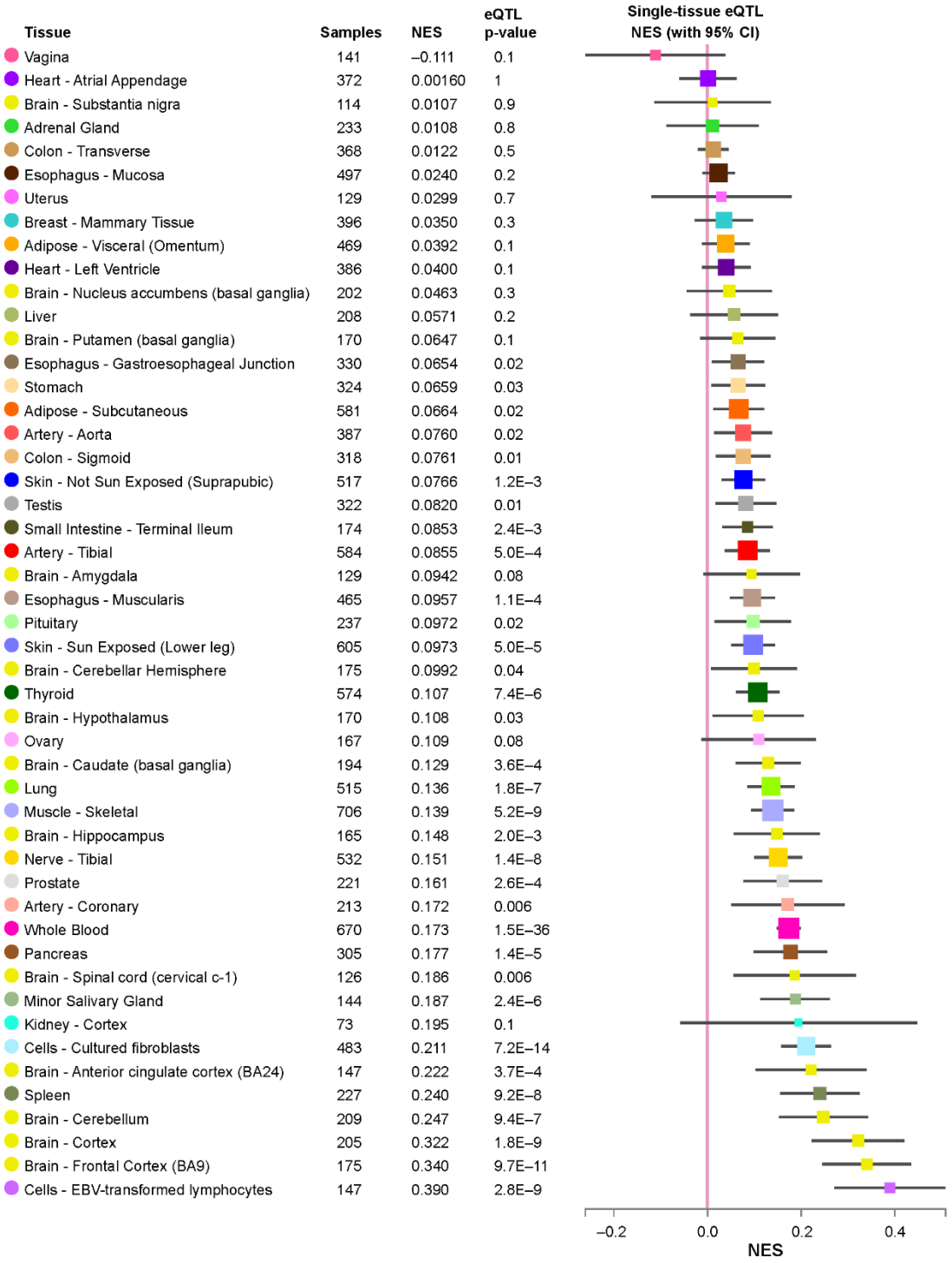
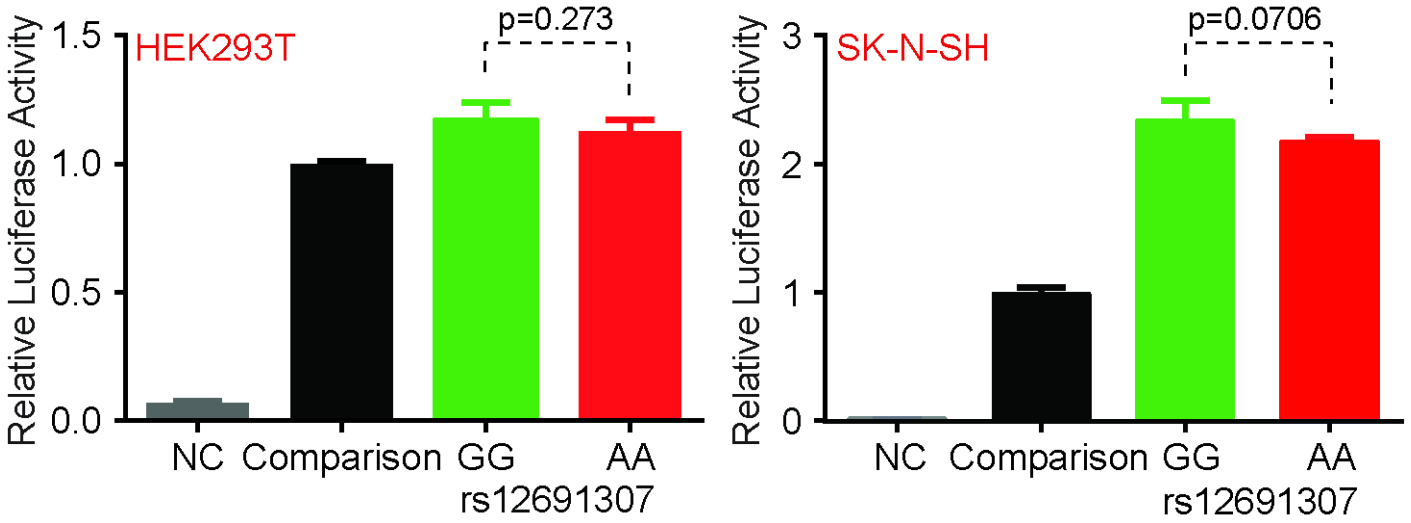
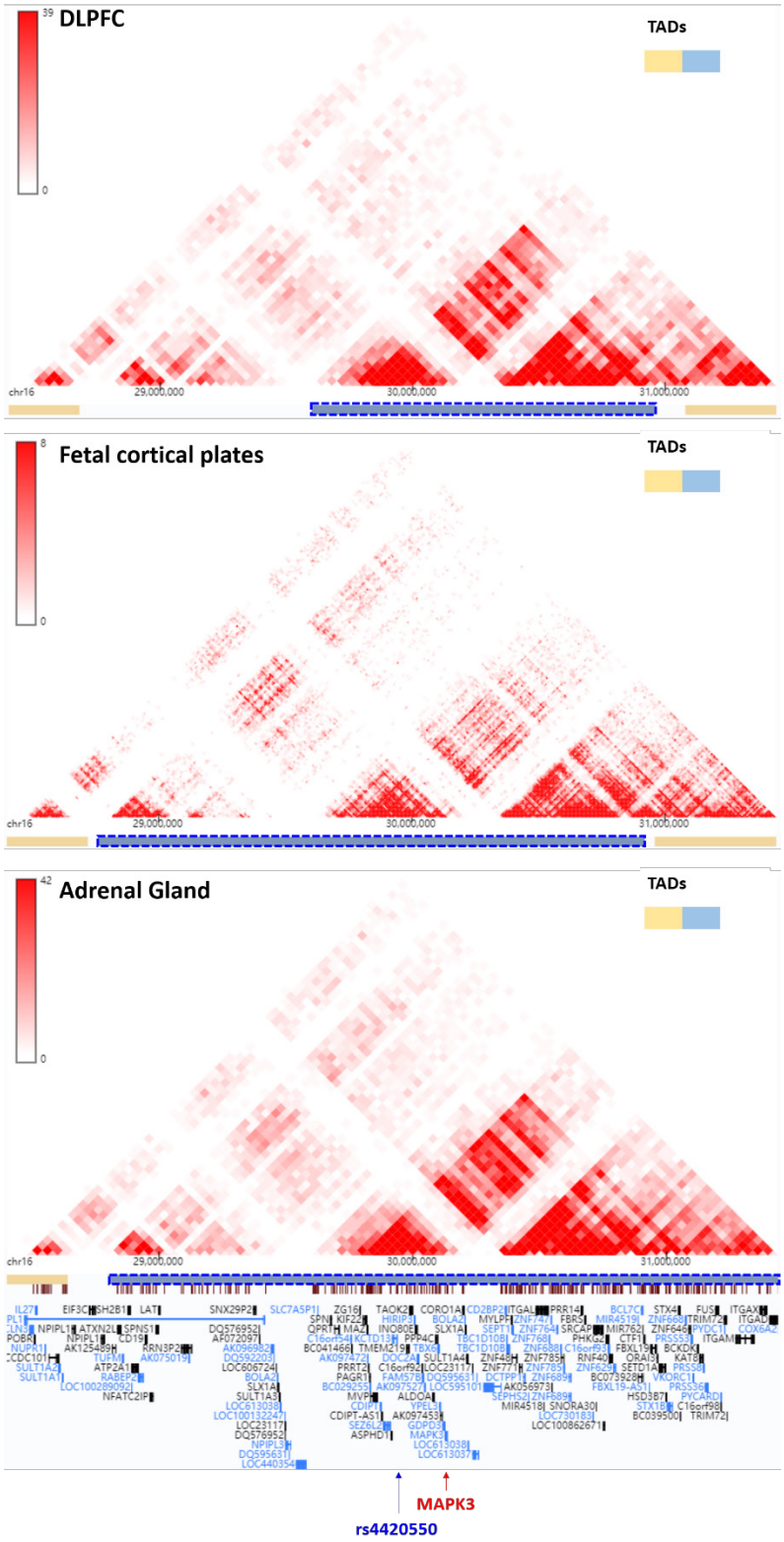


Figure S6. Results of the reporter gene assay testing the regulatory activities of rs12691307. Effects of rs12691307 allele variation on pGL3 promoter activity are shown in the panels for HEK293T and SK-N-SH cells. The "Comparison" in the figure represents the empty pGL3 promoter vector. The Y-axis values represent fold changes of luciferase activity relative to the empty pGL3 promoter vector.



Supplemental Figure S7. rs4420550 and MAPK3 are located in the same topologically associated domain (TAD) in human brain and other tissues according to Hi-C data. Details about the sample collection and Hi-C data processing are described in the original studies (28,29). Visualization was performed using the 3D Genome Browser (http://promoter.bx.psu.edu/hi-c/view.php) (30) in human genome (hg19) with a resolution of 40-kb.



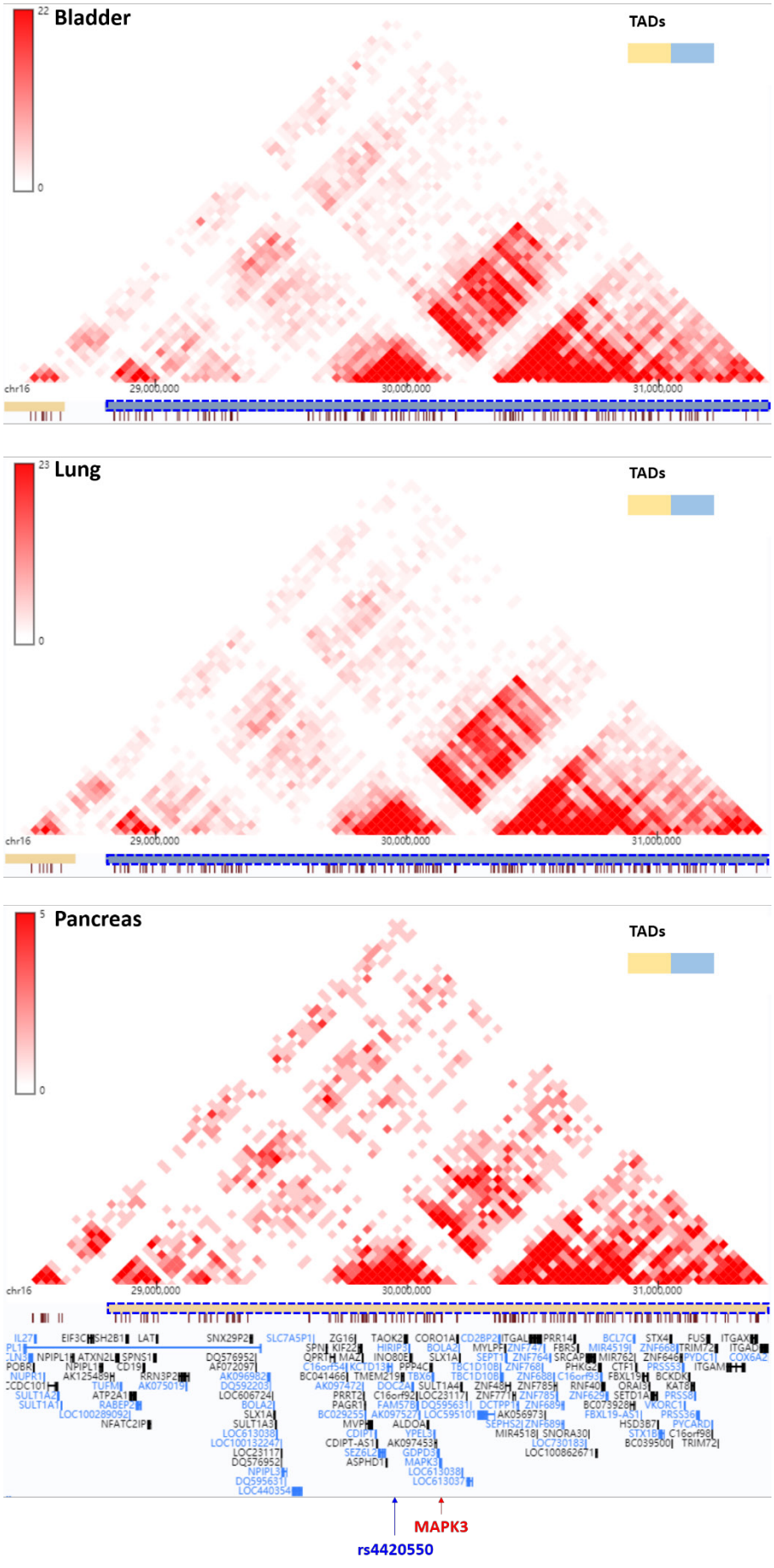

Bladder
22
0
TADs
chr16
29,000,000
30,000,000
31,000,000
Lung
23
0
TADs
chr16
29,000,000
30,000,000
31,000,000
Pancreas
5
0
TADs
chr16
29,000,000
30,000,000
31,000,000
MAPK3
rs4420550

Figure S8. Off-target effect of CRISPR/Cas9 during rs4420550 editing was determined by T7EN1 assay. Out of the top 12 potential off-target sites, typical DNA cleavage of Off-9 site was detected in both sgRNA-Ctrl and sgRNA-rs4420550 transfected cells (A), which were then verified by Sanger sequencing, and no overlap peak was found in PCR products; the off-target site sequence was marked by dotted line (B). These data showed no CRISPR/Cas9 off-target effect in our study.

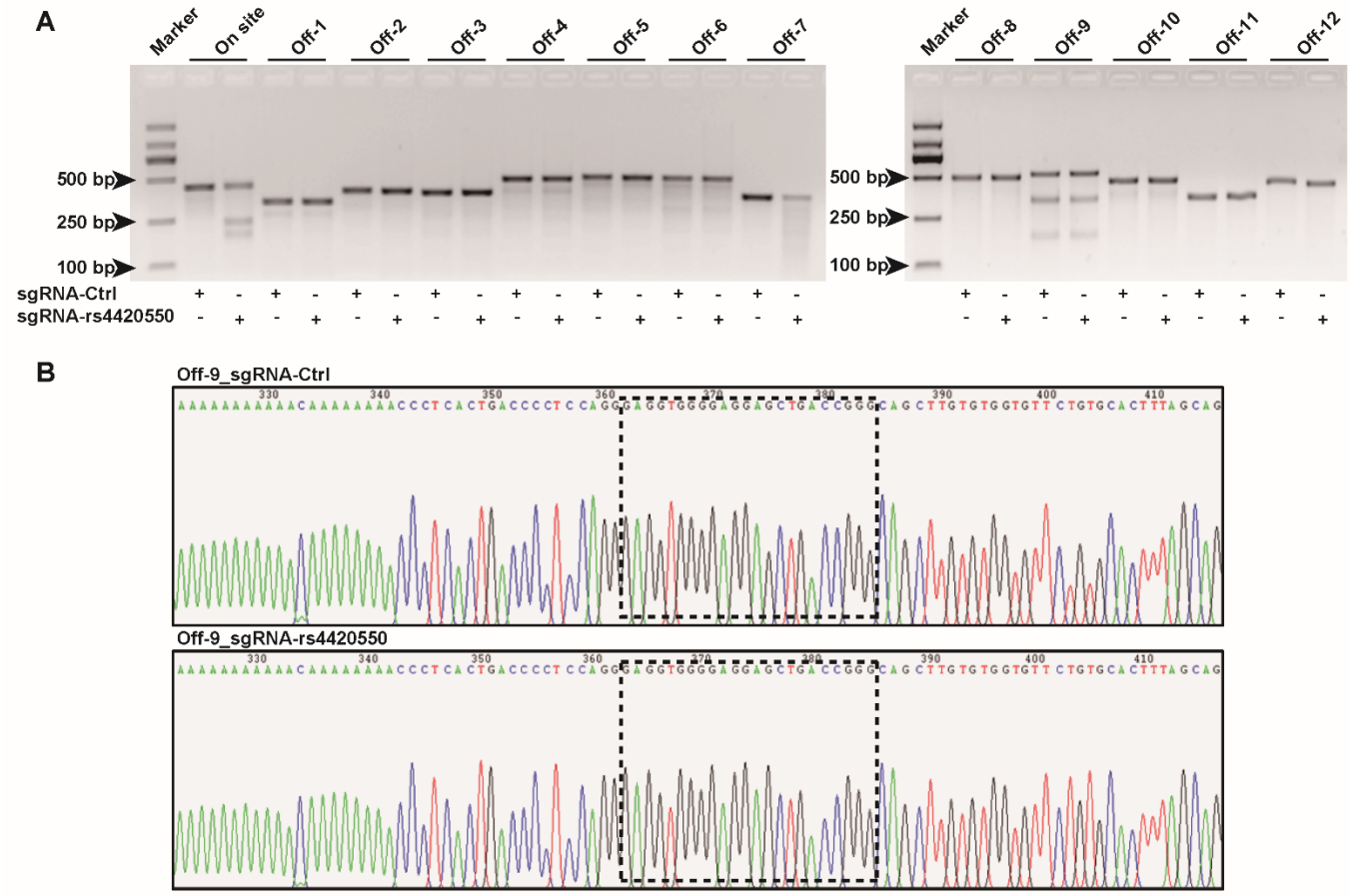


Figure S9. RT-qPCR analysis of genes of interest in rs4420550-edited/unedited HEK293T cells. Expression level of target genes was normalized to the housekeeping gene *RPS13*. The quantitative analyses by qPCR were carried out using the SYBR green mix (Roche, USA) on an ABI Prism 7900 system (Applied Biosystems). Relative expression level is shown as $2^{-\Delta\Delta Ct}$. Each dot represents the average of triplicated quantifications. Error bars depict mean ± s.d.

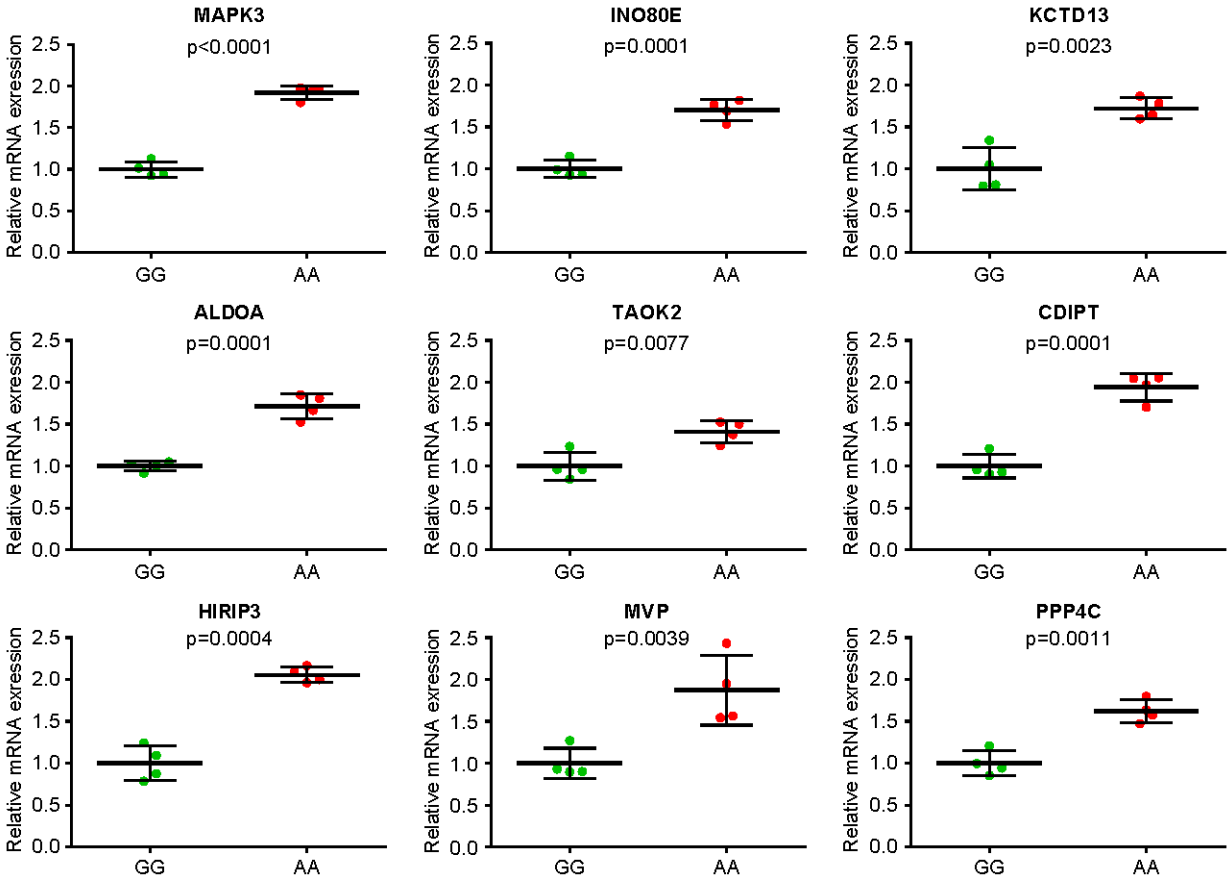


Figure S10. Size distributions of sequenced fragments with pass-QC reads.

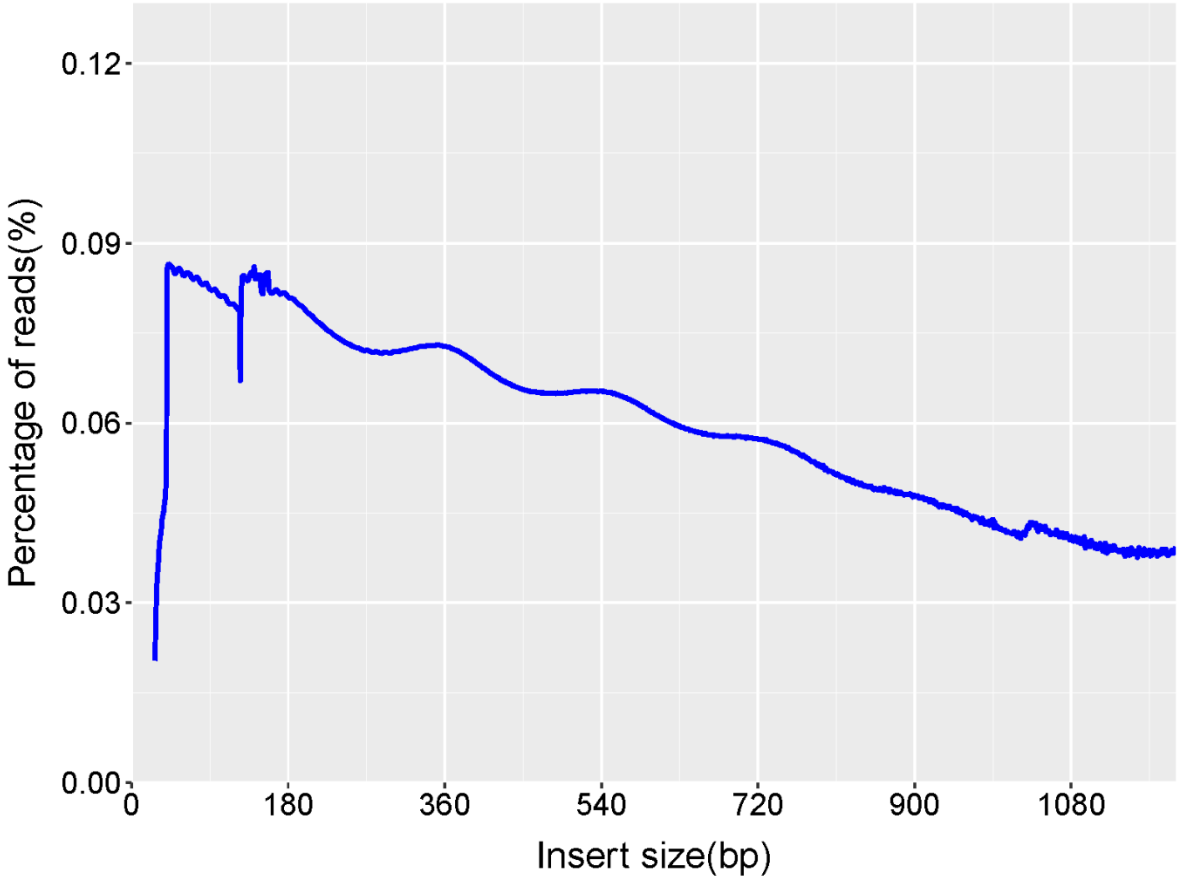


Figure S11. Heatmap illustrates the intensity of signals around transcription start sites (TSS) for each sample. The sequence tag intensity was significantly enriched around TSS in all samples.

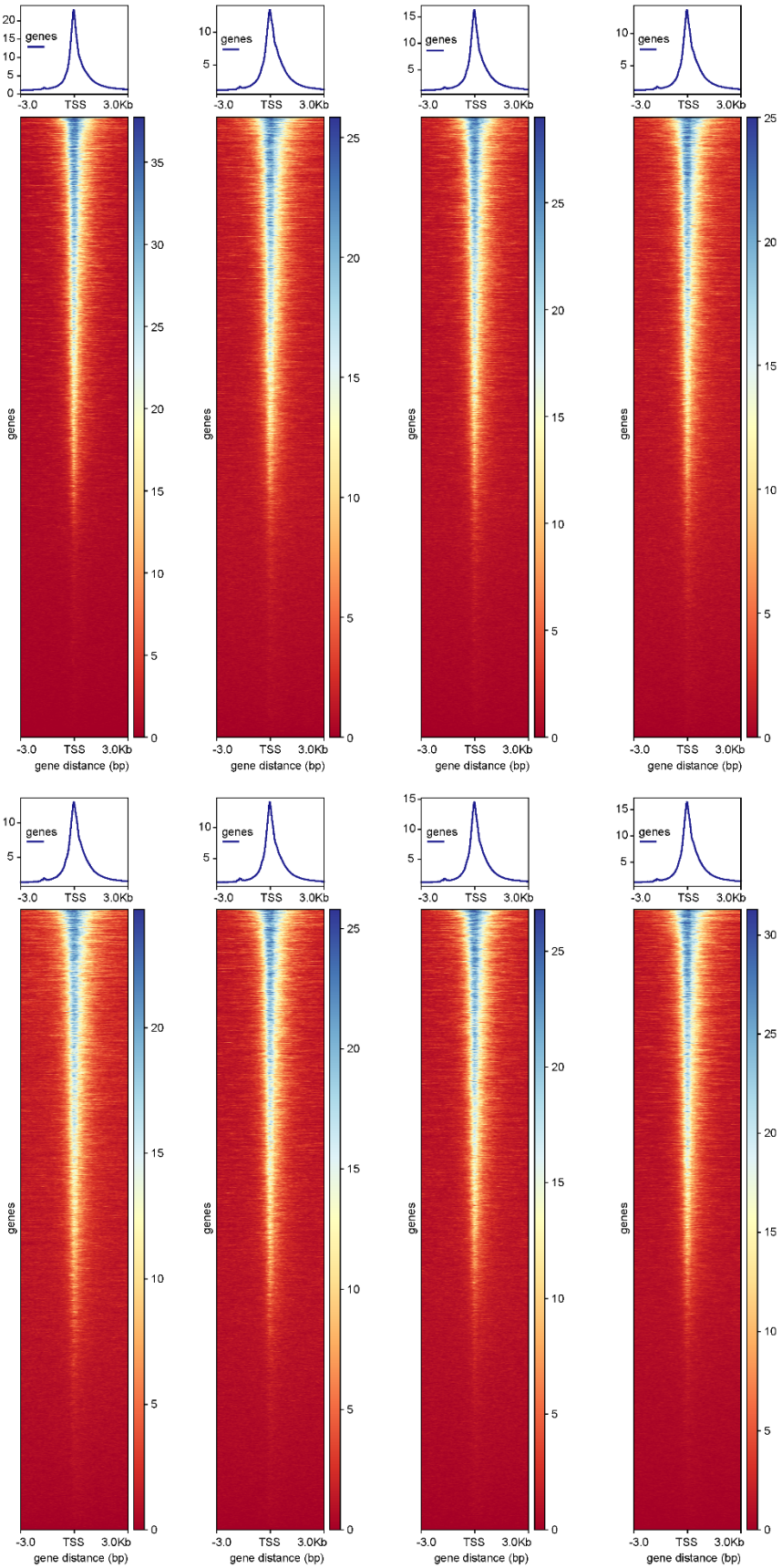


Figure S12. Differences in chromatin accessibility of the *MAZ* TSS regulatory region between the rs4420550-[AA] and [G/G] HEK293T cells. Bins per million mapped reads (BPM) was used to quantify the coverage of peak using bamCoverage (binSize=10) in the deeptools (https://deeptools.readthedocs.io/en/develop/). The ATAC-Seq tag intensity around transcriptional start site (TSS) (±3 kb window) was calculated and plotted using computeMatrix and plotHeatmap programs in the deeptools. The Wald test in DESeq2 was used to identify differently accessible peaks between groups.

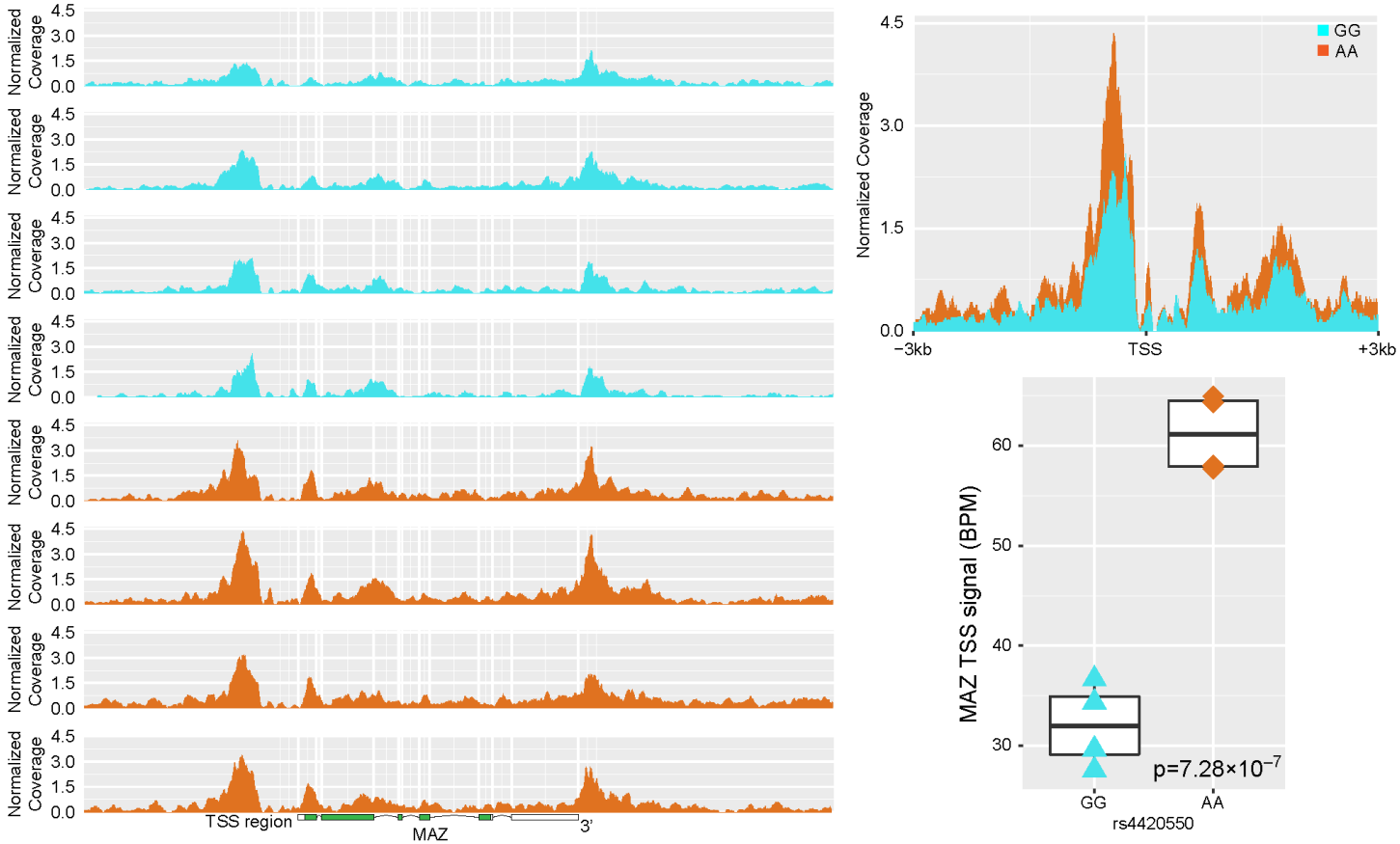


Figure S13. Differences in chromatin accessibility of the *INO80E* TSS regulatory region between the rs4420550-[AA] and [G/G] HEK293T cells. Bins per million mapped reads (BPM) was used to quantify the coverage of peak using bamCoverage (binSize=10) in the deeptools (https://deeptools.readthedocs.io/en/develop/). The ATAC-Seq tag intensity around transcriptional start site (TSS) (±3 kb window) was calculated and plotted using computeMatrix and plotHeatmap programs in the deeptools. The Wald test in DESeq2 was used to identify differently accessible peaks between groups.

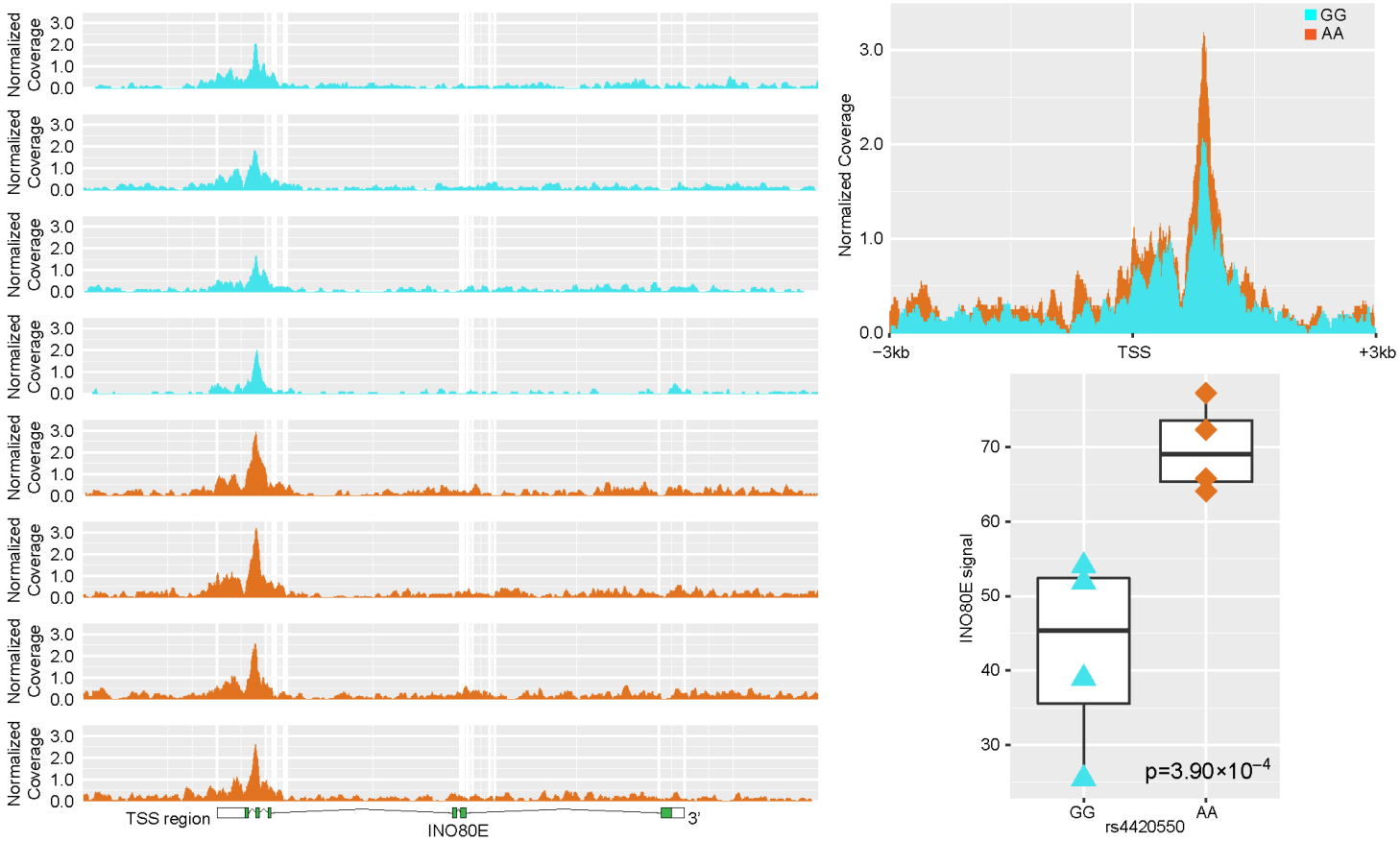


Figure S14. Functional prediction of rs4420550 in binding transcription factors using RegulomeDB (26) and HaploReg v4.1 (27).

RegulomeDB

Motifs Filter:

Method	Location	Motif	Cell Type	PWM	Reference
PWM	chr16:29938816..29938830	Pax-6			16381825
PWM	chr16:29938814..29938827	CREB3L1			23332764

HaploReg v4.1

Regulatory motifs altered

Position Weight Matrix ID (Library from Kheradpour and Kellis, 2013)	Strand	Ref A	Alt G	Match on: Ref: TCCAGAAGCCTTGATGTTGCTAGGAGCTGACCTGGCACCAACAGCACGGCCATGGTGGT Alt: TCCAGAAGCCTTGATGTTGCTAGGAGCTGGCCTGGCACCAACAGCACGGCCATGGTGGT
HDAC2_disc3	-	7.4	-1.6	RSHRCTGKSCRYKGTSCTGA
NRSF_disc3	-	12.5	12.6	GGMGCTGTCC
NRSF_disc4	-	12.3	12.5	SCYSNSCNNSSNSS
Pax-6_2	+	15.6	11.5	CTGWSCTGGAACTM
Sin3Ak-20_disc7	-	13.1	13.5	GSNSCTSNSSNNSS

Figure S15. *MAPK3* mRNA expression in HEK293T cells with CREB3L1/HDAC2/PAX6 knockdown or overexpression. Each transcription factor was knocked down or overexpressed one time, and *MAPK3* mRNA expression was quantified using RT-qPCR.

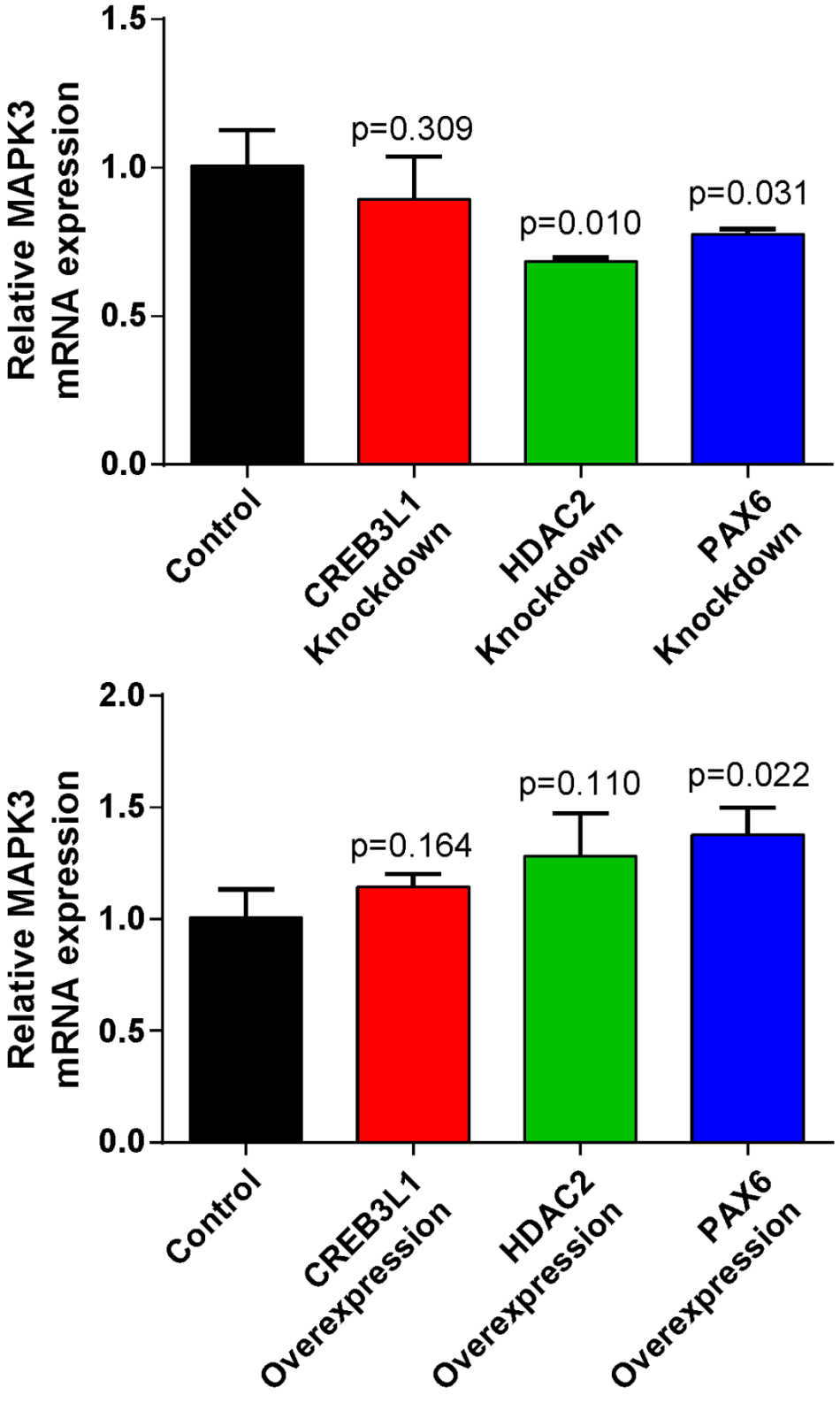


Figure S16. EMSA analysis of DNA sequences spanning rs4420550 (A and G alleles) with nuclear extract from HEK293T cells. Lane 1-2: negative control; Lane 3-4: the probes containing the A or G allele did not bind any HEK293T cell nuclear extracts significantly. Lane 5-6: after overexpressing HDAC2, the probes containing the A or G allele did not bind any HEK293T cell nuclear extracts significantly. Lanes 7-8: after overexpressing PAX6, the probes containing the A or G allele did not bind any HEK293T cell nuclear extracts significantly.

Biotin labeled G-allele	-	+	-	+	-	+	-	+
Biotin labeled A-allele	+	-	+	-	+	-	+	-
Nuclear Extract	-	-	+	+	+	+	+	+
HDAC2 overexpression	No	No	No	No	Yes	Yes	No	No
PAX6 overexpression	No	No	No	No	No	No	Yes	Yes
Lane	1	2	3	4	5	6	7	8

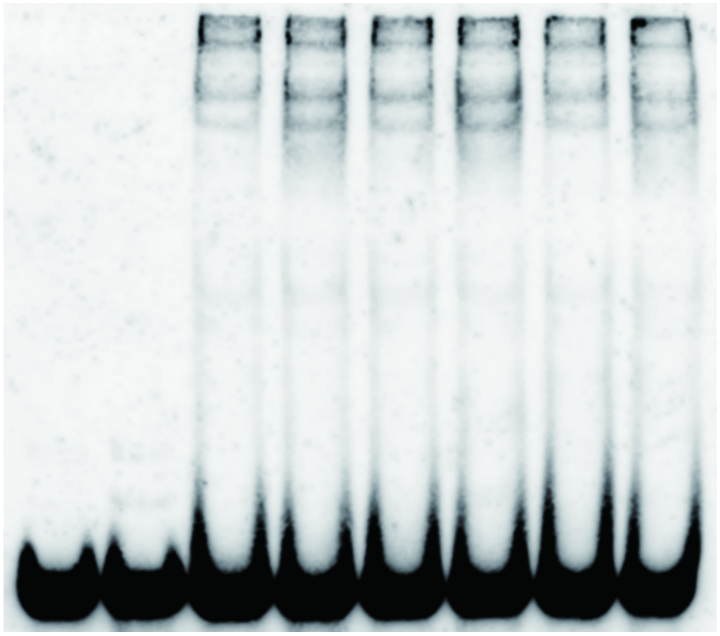


Figure S17. *MAPK3* mRNA expression in rs4420550-[A/A] and [G/G] HEK293T cells with HDAC2 or PAX6 overexpression. *MAPK3* mRNA expression was quantified using RT-qPCR.

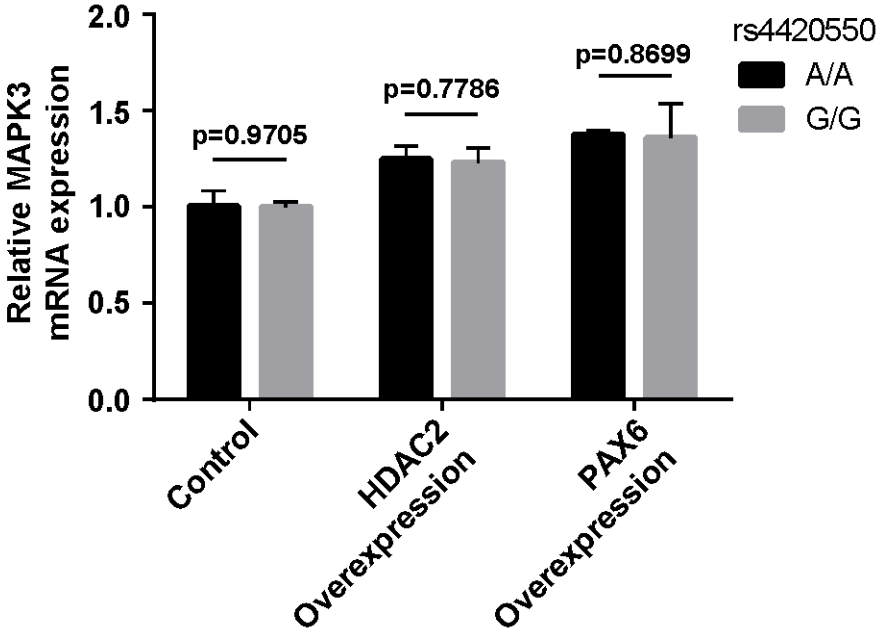
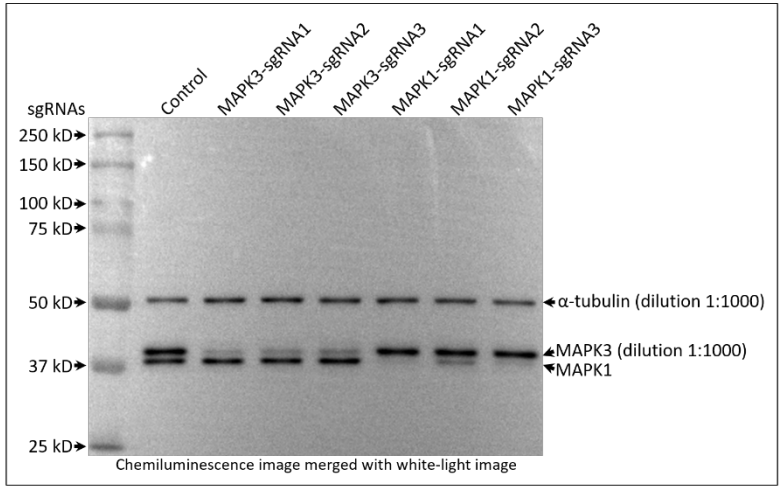
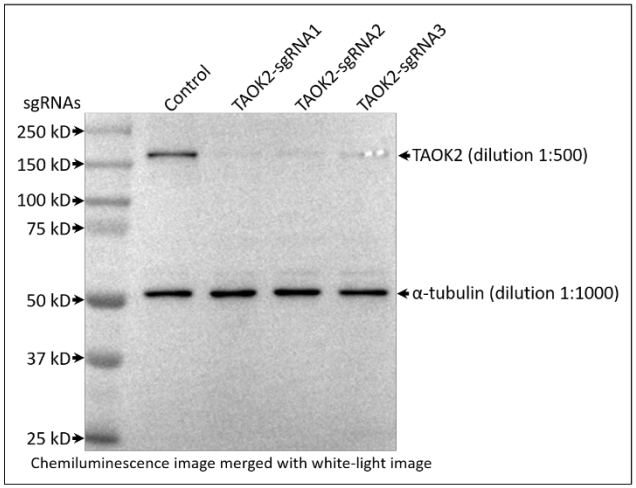
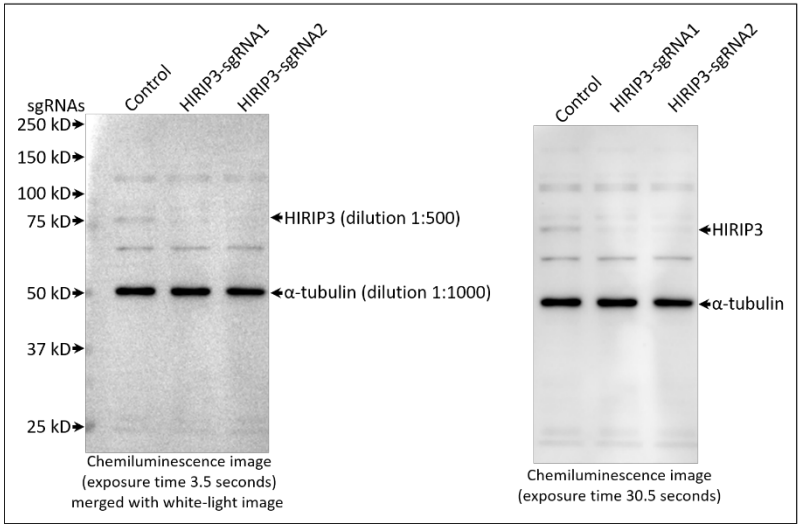


Figure S18. The antibodies against these 16p11.2 proteins have been verified through CRISPR/Cas9 knockdown plasmids.







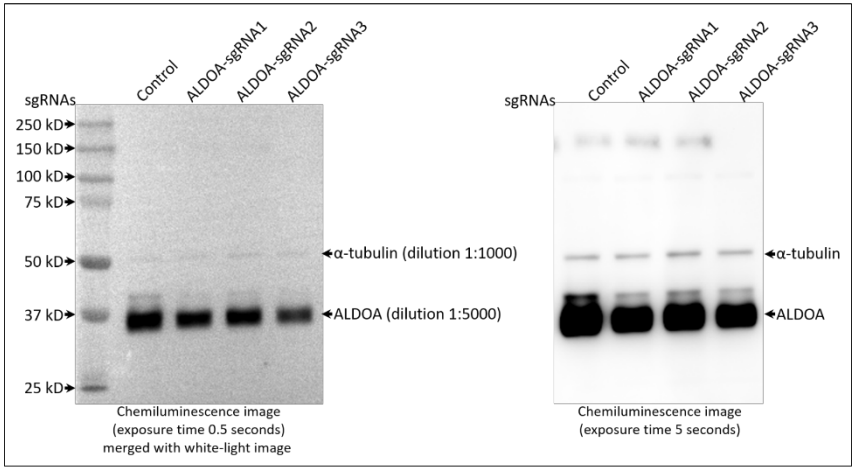
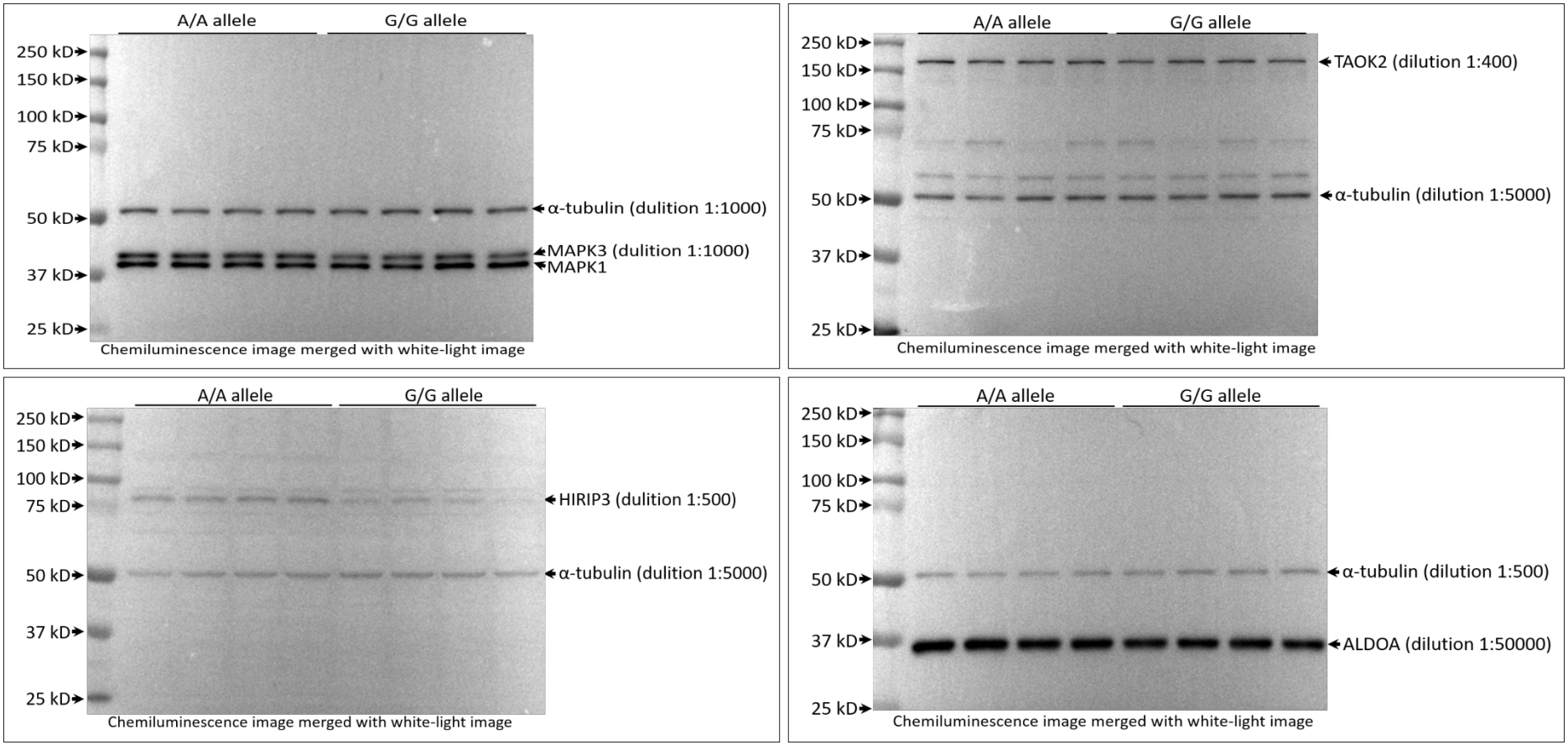


Figure S19. Full-length western blots (merged with prestained protein standards).



Supplemental References

1. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado-Torres L, et al. (2018): Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci*. 21:1117-1125.
2. Ng B, White CC, Klein HU, Sieberts SK, McCabe C, Patrick E, et al. (2017): An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci*. 20:1418-1426.
3. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. (2016): Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci*. 19:1442-1453.
4. GTEx Consortium, Laboratory Data Analysis, Coordinating Center -Analysis Working Group, Statistical Methods groups-Analysis Working Group, Enhancing GTEx groups, NIH Common Fund, et al. (2017): Genetic effects on gene expression across human tissues. *Nature*. 550:204-213.
5. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FCP, et al. (2018): Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 362.
6. PsychEncode Consortium, Akbarian S, Liu C, Knowles JA, Vaccarino FM, Farnham PJ, et al. (2015): The PsychENCODE project. *Nat Neurosci*. 18:1707-1712.
7. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. (2018): Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 362:6420.
8. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. (2016): Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 48:481-487.
9. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. (2018): Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun*. 9:918.
10. Pardinas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. (2018): Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*. 50:381-389.
11. Naumova N, Smith EM, Zhan Y, Dekker J (2012): Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods*. 58:192-203.
12. Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, et al. (2007): Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc*. 2:1722-1733.
13. Gupta RM, Hadaya J, Trehan A, Zekavat SM, Roselli C, Klarin D, et al. (2017): A genetic variant associated with five vascular diseases is a distal regulator of Endothelin-1 gene expression. *Cell*. 170:522-533 e515.
14. Chang H, Yi B, Ma R, Zhang X, Zhao H, Xi Y (2016): CRISPR/cas9, a novel genomic tool to knock down microRNA in vitro and in vivo. *Sci Rep*. 6:22312.
15. Zhao L, Chang H, Zhou DS, Cai J, Fan W, Tang W, et al. (2018): Replicated associations of FADS1, MAD1L1, and a rare variant at 10q26.13 with bipolar disorder in Chinese population. *Transl Psychiatry*. 8:270.
16. Liu W, Yan H, Zhou D, Cai X, Zhang Y, Li S, et al. (2019): The depression GWAS risk

allele predicts smaller cerebellar gray matter volume and reduced SIRT1 mRNA expression in Chinese population. *Transl Psychiatry*. 9:333.

17. Li H, Chang H, Song X, Liu W, Li L, Wang L, et al. (2019): Integrative analyses of major histocompatibility complex loci in the genome-wide association studies of major depressive disorder. *Neuropsychopharmacology*. 44:1552-1561.
18. Bolger AM, Lohse M, Usadel B (2014): Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114-2120.
19. Kim D, Langmead B, Salzberg SL (2015): HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 12:357-360.
20. Anders S, Pyl PT, Huber W (2015): HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 31:166-169.
21. Love MI, Huber W, Anders S (2014): Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.
22. Li H, Zhou DS, Chang H, Wang L, Liu W, Dai SX, et al. (2019): Interactome analyses implicated CAMK2A in the genetic predisposition and pharmacological mechanism of bipolar disorder. *J Psychiatr Res*. 115:165-175.
23. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. (2008): Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 9:R137.
24. Liao Y, Smyth GK, Shi W (2014): featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 30:923-930.
25. Barrett JC, Fry B, Maller J, Daly MJ (2005): Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 21:263-265.
26. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. (2012): Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 22:1790-1797.
27. Ward LD, Kellis M (2012): HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 40:D930-934.
28. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. (2016): A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep*. 17:2042-2059.
29. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, et al. (2016): Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 538:523-527.
30. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. (2018): The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol*. 19:151.