Supplementary Tables

Abbreviation	Trait/disease	Reference (doi)
ADHD	Attention deficit hyperactivity disorder	10.1016/j.jaac.2010.06.008
ALS	Amyotrophic lateral sclerosis	10.1016/j.neuron.2018.02.027
PD	Parkinson's disease	10.1016/S1474-4422(19)30320-5
RA	Rheumatoid arthritis	10.1038/nature12873
BMI	Body mass index	10.1038/nature14177
EduYear	Educational years	10.1038/nature17671
HEIGHT	Height	10.1038/ng.3097
CD	Crohn's disease	10.1038/ng.3359
IBD	Inflammatory bowel disease	10.1038/ng.3359
UC	Ulcerative colitis	10.1038/ng.3359
CAD	Cardiovascular disease	10.1038/ng.3396
DS	Depression symptom	10.1038/ng.3552
Neu	Neuroticism	10.1038/ng.3552
SLE	Systemic lupus erythematosus	10.1038/ng.3603
T2D	Type 2 Diabetes	10.1038/s41588-018-0241-6
DRINKING	Alcohol Assumption	10.1038/s41588-018-0309-3
AD	Alzheimer's disease	10.1038/s41588-018-0311-9
ASD	Autism spectrum disorder	10.1038/s41588-019-0344-8
BD	Bipolar disorder	10.1038/s41588-019-0397-8
MDD	Major depression disorder	10.1038/s41593-018-0326-7
SZ	Schizophrenia	10.1101/2020.09.12.20192922
MS	Multiple sclerosis	10.1126/science.aav7188

Supplementary Table 1. Trait/Disease, abbreviation and reference for GWAS included in LD-score regression analysis

Supplementary Notes

Permutation to evaluate the empirical performance of CLPP cutoff

CLPP was first coined by Hormozdiari, et al. [1], measuring a joint posterior probability from colocalization, and is calculated as the product of posterior inclusion probability in two traits. A 1% cutoff is used in the original paper, and then widely applied by others [2–4]. We performed two permutation analyses on our real data to test the empirical performance of the 1% CLPP cutoff. First, we considered a scenario where there are no true eQTLs. Using expression of 2,457 genes on chr1 from the PsychENCODE data, we randomly permuted the sample labels and then performed eQTL analysis followed by eCAVIAR colocalization analysis with PGC2 Schizophrenia GWAS summary statistics. We found that there is no SNP with a CLPP > 1%. Next we considered the scenario where there are true eQTLs in the dataset, but their genome locations are independent of the GWAS associations. We extracted eQTL results for the 920 genes on chr1 that have a lead eQTL variant with p < 1e-6, and performed statistical fine-mapping on each gene. The posterior inclusion probabilities were then shuffled and used for colocalization analysis with eCAVIAR using the same Schizophrenia GWAS data. Only a single SNP passed the 1% CLPP cutoff in this analysis.

Validation of rs72986630 effect in chromatin accessibility and gene expression data

To further investigate one prioritized functional variant rs72986630 that reside in REST TF binding site overlapping TSS of *ZNF823*, we queried our unpublished ATAC-seq data set (Bendl et al., in preparation) of neuronal and non-neuronal samples from ACC brain region generated on postmortem human brains from CommonMind cohort [5]. This dataset consists of samples from 370 donors (114 SZ cases, 64 BD cases, 64 controls) with rs72986630 MAF of 6.0%. Since only two donors carry the ALT/ALT (i.e. T/T) genotype, we excluded them for further analysis.

To generate ATAC-seq data set, neuronal and non-neuronal cell populations were isolated from postmortem tissue by fluorescence-activated nuclear sorting using anti-NeuN antibody. ATAC-seq libraries were created using an established protocol [6]. Raw sequencing reads were mapped to human genome hg38 using STAR aligner [7]. The samples of the same cell type (neuron / non-neuron) and genotype at rs72986630 (CC / CT) were subsampled and merged, creating 4 BAM files with a uniform depth of 1 billion pair-end reads. Subsampling ratios were calculated per each sample individually within those four respective groups (genotype by cell type) to ensure that each of them contributed the same number of reads, regardless of their per-sample read counts. Using these BAM files, bigWig files were created and peaks were called by the MACS (v2.1) with the same parameters as described in Hauberg, et al.[48], but using an FDR threshold of 0.01. After removing peaks overlapping the blacklisted genomic regions and peaks not being sufficiently accessible (CPM>1 in at least 10% of samples was required), 498,183 peaks remained. The final read count matrix of 664 samples by 498,183 peaks was normalized using the trimmed mean of M-values (TMM) method. The following covariates were selected by Bayesian Information

Criterion (BIC) method to be added to the base covariates, i.e. genotype by cell type: mean GC content, fraction of reads with GC content 0-19%, 20-39%, 40-59%, fraction of reads in peaks, fraction of unmapped reads, AT dropout, and mean insert size. Since our dataset contains up to two samples per individual, we ran differential analysis to get differentially accessible peaks between CC and CT carriers using dream (v1.17.9) [8] that accounts for correlation structure in repeated measures. As an alternative approach, instead of quantifying changes between all open chromatin regions, we performed differential analysis between CC and CT carriers on TF binding sites of REST motif. We used footprinting to narrow down our focus only to 31,534 REST TF binding sites that are bound in at least one set of samples (out of 4 sets, i.e. genotype by cell type) as predicted by TOBIAS[9].

The analysis of differential gene expression between REF/REF (C/C) and REF/ALT (C/T) genotype at rs72986630 followed the same approach as the analysis of chromatin accessibility. We used a subset of 338 homogenate RNA-seq samples of ACC brain region from CommonMind Consortium [5] that originate from the same donors as ATAC-seq samples. We performed differential analysis only for sufficiently expressed protein-coding genes (CPM>1 in at least 30% of samples was required), i.e. we start the analysis with a count matrix of 338 samples by 14,893 genes that were normalized by trimmed mean of M-values (TMM) method. The following technical covariates were selected by BIC method: institution, expression profiling efficiency, intronic rate, intragenic rate, fraction of reads with GC content 20-39%, 40-59%, and AT dropout.

1. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
2. Klarin, D. *et al.* Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat. Genet.* **51**, 1574–1579 (2019).
3. Bonder, M. J. *et al.* Identification of rare and common regulatory variants in pluripotent cells using population-scale transcriptomics. *Nat. Genet.* **53**, (2021).
4. Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006-3021.e17 (2021).
5. Hoffman, G. E. *et al.* CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. *Sci. Data* **6**, 180 (2019).
6. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9 (2015).
7. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

8. Hoffman, G. E. & Roussos, P. dream: Powerful differential expression analysis for repeated measures designs. *Bioinformatics*. 37, 192-201 (2021).
9. Bentsen, M. *et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).