bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Title: Transcribed enhancers in the human brain identify novel disease risk mechanisms

Authors: Pengfei Dong[1,2,3,4,5], Gabriel E. Hoffman[1,3,5], Pasha Apontes[1,2,3,4,5,9], Jaroslav Bendl[1,2,3,4,5], Samir Rahman[1,2,3,4,5], Michael B. Fernando[1,2,6,7], Biao Zeng[1,2,3,4,5], James M. Vicari[1,2,3,4,5], Wen Zhang[1,2,3,4,5], Kiran Girdhar[1,2,3,4,5], Kayla G. Townsley[1,2,6], Ruth Misir[1,2,3,4,5], the CommonMind Consortium, Kristen J. Brennand[1,2,4,6,7,10], Vahram Haroutunian[2,4,6,9], Georgios Voloudakis[1,2,4], John F. Fullard[1,2,3,4,5], Panos Roussos[1,2,3,4,5,8,9,11]

Affiliations:

[1]Pamela Sklar Division of Psychiatric Genomics
[2]Friedman Brain Institute,
[3]Icahn Institute for Data Science and Genomic Technology,
[4]Department of Psychiatry,
[5]Department of Genetics and Genomic Science
[6]Department of Neuroscience
[7]Black Family Stem Cell Institute
Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[8]Center for Dementia Research, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA
[9]Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, New York, 10468, USA
[10]Current address: Department of Psychiatry, Yale University, New Haven, CT 06511, USA
[11]Lead Contact: panagiotis.roussos@mssm.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Summary

Enhancer RNAs (eRNAs) constitute an important tissue- and cell-type-specific layer of the regulome. Identification of risk variants for neuropsychiatric diseases within enhancers underscores the importance of understanding the population-level variation of eRNAs in the human brain. We jointly analyzed cell type-specific transcriptome and regulome data to identify 30,795 neuronal and 23,265 non-neuronal eRNAs, expanding the catalog of known human brain eRNAs by an order of magnitude. Examination of the population-level variation of the transcriptome and regulome in 1,382 brain samples identified reproducible changes affecting *cis*- and *trans*-co-regulation of eRNA-gene modules in schizophrenia. We show that 13% of schizophrenia heritability is jointly mediated in *cis* by brain gene and eRNA expression. Inclusion of eRNAs in transcriptome-wide association studies facilitated fine-mapping and functional interpretation of disease loci. Overall, our study characterizes the eRNA-gene regulome and genetic mechanisms in the human cortex in both healthy and disease states.

Keywords: enhancer RNA, neuropsychiatric disorders, enhancer expression QTL, GWAS, fine-mapping, gene regulation

Highlights

- Catalog and characterize cell-type-specific transcribed enhancers in the human cerebral cortex.
- *Cis*- and *trans-eRNA-gene* co-expression networks dissect cell-type-specific dysregulations of regulome in schizophrenia.
- Adding eRNA to the analysis of gene expression data increases the explained heritability mediated in *cis*-regulation of expression by more than 50% when compared to genes alone.
- Joint eRNA-gene transcriptome-wide association study facilitates functional characterization of schizophrenia risk loci.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Enhancers are key regulatory regions of DNA that exert control over target gene expression from a distance (Long et al., 2016; Schoenfelder and Fraser, 2019). Besides their critical function in orchestrating cell-lineage commitment and development (Heinz et al., 2015; Long et al., 2016), enhancers play pivotal roles in mediating neuronal plasticity and memory formation in the brain (Yap and Greenberg, 2018), and human evolved enhancers are hypothesized to drive advanced cognition (Reilly et al., 2015; Varki et al., 2008; Won et al., 2019). Upon activation, enhancers undergo nucleosomal depletion to expose motifs that facilitate the binding of transcription factors (TFs). TFs and co-factors binding to DNA enhancer regions recruit RNA polymerase II, leading to the covalent modification of flanking histone tails (including H3K27ac and H3K4me3) and to the bi-directional transcription of enhancer RNA (eRNA) (Heinz et al., 2015; Krijger and de Laat, 2016; Rothschild and Basu, 2017). As an emerging class of non-coding RNA, eRNA is widespread at all active enhancers in quantities that are proportional to enhancer activity (Core et al., 2014; Tippens et al., 2020). Rather than merely being by-products of enhancer activity, several lines of evidence suggest that some, if not all, eRNAs are functional (Catarino and Stark, 2018; Gu et al., 2018; Hou and Kraus). More traditional approaches for identifying enhancers have their drawbacks: in contrast to eRNAs, chromatin accessibility also captures non-active enhancers (Thurman et al., 2012) and histone modification marks, such H3K27ac, can capture ‘‘stretch-enhancers’’ or ‘‘super-enhancers’’ of more than 10kb in size (Hnisz et al., 2013; Parker et al., 2013), making it difficult to pinpoint functional loci. On the other hand, enhancers identified via eRNA are more likely to be validated in functional assays (Andersson et al., 2014), possibly due to the activity-dependent expression and smaller size of eRNAs. Although the Functional Annotation of the Mammalian Genome (FANTOM) project successfully annotated ~65,000 eRNAs across more than 400 human tissues and cell types using only transcriptomic signatures captured by cap analysis of gene expression (CAGE-seq) (Andersson et al., 2014), it is thought that the majority of eRNAs have yet to be identified (Core et al., 2014; Tippens et al., 2020). In the human brain, the annotation includes only a few thousand eRNAs and does not consider their distribution across the different cell types (Andersson et al., 2014). As such, a systematic cell-type-specific map of human brain eRNA expression would be an important step towards developing a more thorough understanding of enhancer functional units in the brain and their contribution to disease.

The majority of common neuropsychiatric disease-associated variants lie in the non-coding genome (Grove et al., 2019; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Stahl et al., 2019; Sullivan and Geschwind, 2019; Visscher et al., 2017). Brain-associated enhancers are overrepresented both for risk alleles and heritability of neuropsychiatric traits (Finucane et al., 2015; Wang et al., 2018a), highlighting their critical role in disease. However, a thorough analysis of enhancer function and their involvement in the etiology of neuropsychiatric disease is hampered by limited access to biospecimens and available molecular assays. As eRNA expression levels reflect enhancer activity such an approach provides an alternative means to quantitatively investigate regulatory circuitry (Andersson et al., 2014; Chen et al., 2018; Hauberg et al., 2019; Kim et al., 2010; Mikhaylichenko et al., 2018; Murakawa et al., 2016). Although eRNA molecules are often not polyadenylated and, hence, are less stable, they can be captured by deeply sequencing total RNA (Chen et al., 2018; Hauberg et al., 2019; Kim et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2010). Issues concerning the inherent instability and low expression rates of eRNAs (Sartorelli and Lauberth, 2020) can be ameliorated when quantification is applied to large cohorts, allowing for the identification of robustly expressed eRNAs across multiple replicates.

To further our understanding of the roles played by enhancer elements in neuropsychiatric disorders, we generated comprehensive multi-omics reference maps for both neuronal and non-neuronal cells and used this to develop a computational scheme that accurately cataloged 30,795 neuronal and 23,265 non-neuronal eRNAs in the human cerebral cortex. We then examined the population-level variation of eRNA expression by analyzing 1382 RNA-seq libraries from 332 schizophrenia (SCZ) and 442 Control postmortem brains from the CommonMind Consortium (CMC). We found eRNAs that are robustly expressed in cortical tissue, explored eRNA-gene transcriptional coordination, and dissected cell-type-specific SCZ-associated regulatory circuits, as well as the TFs that drive them. In addition, we examined SCZ heritability shared with expression quantitative trait loci (eQTL) for genes and eRNAs and performed transcriptome-wide association analysis to facilitate fine-mapping and functional interpretation of disease loci. Collectively, our study characterizes the population-level variation of eRNA-gene regulome in the human cortex in both healthy and disease states.

Results

Comprehensive multi-omics maps of neuronal and non-neuronal cells

To annotate and characterize the expression pattern of eRNAs in the human cortex, we first performed multi-omics profiling, including ribosomal-RNA (rRNA) depleted total-RNA-seq, ATAC-seq, and ChIP-seq for H3K4me3 and H3K27ac) in neuronal (NeuN+) and non-neuronal (NeuN-) nuclei, isolated by fluorescence-activated nuclear sorting (FANS), from five brain regions (Brodmann areas 10, 17, 22, 36, and 44) of 10 control individuals (**Figure 1A and 1B**). After extensive quality control, including assessing cell type, sex, and genotype concordance (Methods), a total of 14.5 billion uniquely mapped paired-end read pairs for RNA-seq (N=93), 4.3 billion for ATAC-seq (N=98), 11.0 billion for H3K4me3 ChIP-seq (N=96), and 9.6 billion for H3K27ac ChIP-seq (N=96) were obtained. Additionally, in a subset of samples (N=6), we mapped repressors and long-range chromatin interactions in neurons and non-neurons by performing H3K27me3 ChIP-seq and Hi-C, respectively.

To validate the cell-type specificity of our data, we performed a transcriptome deconvolution analysis with reference markers from single-cell analysis (Lake et al., 2018). The neuronal samples were strongly enriched for glutamatergic and GABAergic neurons, while the non-neuronal cells were enriched for oligodendrocytes, followed by astrocytes and microglia (**Figure S1A**), thus confirming the cell-type specificity of FANS. In addition, our ChIP-seq and ATAC-seq data were highly concordant with previously published reports (**Figure S1B**). Thus, we generated high-quality, deeply sequenced, cell type-specific transcriptomic and epigenomic maps of the human cortex.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A catalog of cell type-specific enhancers in the human cortex

eRNA was previously defined by the FANTOM project based on density enrichment of bidirectional CAGE tags (Andersson et al., 2014). We sought to utilize our multi-omics dataset to further expand the catalog of eRNAs in the human brain. As an exploratory step, we compared the molecular profiles between expressed enhancers and those that were not expressed. To define a set of expressed enhancers, we collected FANTOM5 eRNAs that overlapped with our H3K27ac peaks (FANTOM5 eRNA, FEs). Additionally, we also included noncoding open chromatin regions (OCRs) that were missed by FANTOM5 but had bidirectional CAGE tags (Extra expressed, EEs). OCRs that did not overlap with FANTOM5 eRNA and had no CAGE tags were defined as non-expressed enhancers (Non-expressed, NEs) (**Figure 1C**). We reasoned that the eRNA transcription initiation sites correspond to TF binding sites (Core et al., 2014; Tippens et al., 2020), which can be determined by the ATAC-seq peak summit. Indeed, the FE positions, which were determined by CAGE tags, captured the ATAC-seq signal summit and were flanked by well-positioned nucleosomes. Its position relative to accessible chromatin and nucleosomes aligned well with the OCR-derived EEs and NEs, suggesting OCR summits pinpoint eRNA positions (**Figure 1C**). Both FEs and EEs exhibit local transcription signals, as well as typical active enhancer chromatin modifications, including H3K4me3 and H3K27ac enrichment (**Figure 1C**). It's worth noting that, compared to FEs, EEs had markedly lower levels of H3K4me3 signal, suggesting that FANTOM5 eRNA might be biased towards enhancers with high levels of H3K4me3. In contrast, NEs were depleted of such active enhancer histone marks and, in turn, displayed much lower expression levels.

Based on the distinct epigenomic signatures, as well as accurate enhancer positioning, we developed a supervised machine learning scheme to expand and capture the cell-type-specific regulome of eRNAs in the human brain (**Figure 1D**). For each cell type, we used the central and flanking multi-omics signals, as well as the genomic annotation of the OCRs, as input for random forest models to select features differentiating expressed enhancers (positive sets, FEs, and EEs) from non-expressed enhancers (negative sets, NEs) (Methods). The resulting models are of high accuracy, as measured by the area under the receiver operating characteristic (0.97 and 0.95 for neuron and non-neuron, respectively) and the area under the precision-recall curves (0.97 and 0.93 for neuron and non-neuron, respectively) (**Figure 1E** and **Figure 1F**). Interestingly, for both cell types, antisense and sense RNA-seq signal at the upstream and downstream flanking region, respectively, ranked as the most important feature to differentiate expressed from non-expressed enhancers (**Figure S1D**). Overall, we identified 36,927 neuronal and 27,379 non-neuronal eRNAs, among which only 2,487 (6.73%) of neuronal and 2,833 (10.3%) of non-neuronal eRNAs are within 2kb of FANTOM5 defined eRNAs. In sharp contrast to the background (i.e. non-expressed non-coding OCRs), the identified eRNAs exhibited strong active enhancer signatures (H3K4me3 and H3K27ac peaks) as well as long-range, chromatin interactions (**Figure 1G**). Specifically, almost all (94% non-neuronal and 97.4% neuronal) eRNAs overlapped with H3K27ac peaks (odds ratio, OR=43.1 non-neuronal, OR=119.3 neuron, $p<10^{-16}$ for both, Fisher's test) and eRNAs were overrepresented in super-enhancer (SE) regions. On average, a single SE contained 7.47 and 6.50 eRNAs in neuronal and non-neuronal cells, respectively (**Figure S1E**) and, overall, 86% of neuronal and 91% of non-neuronal SEs expressed at least one eRNA. In addition, eRNAs had a much higher density in SE regions ($p<10^{-16}$ for both, Wilcoxon test) (**Figure S1F**). To examine

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

if the identified eRNAs represent distal transcription initiation, we superimposed strand-specific CAGE tags on the identified eRNAs. This approach provides a means to capture TSS, rather than elongation signals (Andersson et al., 2014; Yao et al., 2015). As expected, both intergenic and intronic eRNAs exhibited bi-directional CAGE tags, confirming that the identified eRNA loci represent bi-directional TSSs (**Figure 1H**).

We then quantified the expression of both genes and eRNAs. We found that 30,795 (83.4%) neuronal and 23,265 (85.0%) non-neuronal eRNAs were expressed at >0.25 counts per million in >10% of the samples used in the current study. We subsequently compared the differences in epigenome profiles between expressed eRNAs and expressed gene promoters (**Figure S1G**). As expected, eRNAs possessed a distinct epigenomic signature compared to typical promoters. Both protein-coding gene and long intergenic noncoding RNA (lincRNA) promoters displayed abundant H3K4me3 and H3K27ac signals, whereas eRNA loci were enriched for H3K27ac but had much lower levels of H3K4me3. Moreover, eRNA loci also had lower H3K27me3 levels compared to the protein-coding genes and lincRNAs. Collectively, we successfully annotated 30,795 neuronal and 23,265 non-neuronal eRNAs (**Table S1**), greatly expanding the repertoire of known human brain transcribed enhancers. Altogether, these eRNA loci exhibit active enhancer signatures, represent transcription initiation, and are distinguishable from typical promoters.

eRNA expression captures cell-type-specific enhancer function

To further characterize neuronal and non-neuronal eRNAs, we annotated chromatin states by jointly analyzing ChIP-seq data using ChromHMM (**Figure S2A and S2B**) (Ernst and Kellis, 2017). Compared to other *cis*-regulatory elements (CREs), such as promoters and polycomb repressors, enhancers were markedly different between the two cell types (**Figure S2C**). In line with this, the identified eRNAs were also largely non-overlapping, confirming the strong cell-type specificity of enhancer elements (**Figure S2D**). We then quantified cell type differences by modeling the read count matrices of gene/eRNA, and CREs from ATAC-seq and ChIP-seq, as well as confounders selected by covariate analysis (Methods and **Table S2**). Despite their modest expression levels, ~90% of eRNAs were differentially expressed (DE) between the two cell types (**Figure S2E**). In addition, we assessed the variance explained by different factors (Methods), and cell type was the strongest source of variation for all assays (**Figure S2F**). The DE analysis identified 8,864 genes and 22,669 eRNAs that were upregulated in neurons, and 9,140 genes and 22,582 eRNAs upregulated in non-neurons. The effect size of cell-type differences between eRNA expression and enhancer activities, as determined by ChIP-seq or ATAC-seq, were highly correlated for both intergenic and intronic eRNAs (**Figure 2A** and **Figure S2G**), highlighting the observation that eRNA expression quantified by RNA-seq gives a good representation of enhancer activity.

We conducted transcription factor (TF) motif enrichment analysis with the DE eRNAs to detect lineage-specific TFs in neuronal and non-neuronal cells. DE eRNAs were enriched for TFs known to play roles in the corresponding cell function (**Figure 2B**). For example, neuronal eRNAs are overrepresented for key regulators of neuronal development and cognition, such as the NEUROD family and EGR family of TFs (Guillemot, 2007; Poirier et al., 2008). Similarly, non-neuronal eRNAs were enriched for TFs that are known to regulate glial function, including the Forkhead box (FOX) family (Golson and Kaestner, 2016), and SOX12 (Bhattaram et al., 2010; Hoser et al., 2008). In addition, we performed motif enrichment analysis with ATAC-seq data using ChromVar

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(Schep et al., 2017) (**Figure 2B**) and found that enrichment was highly concordant (Spearman correlation coefficient is 0.641 and 0.564 for neurons and non-neurons respectively; $p<10^{-16}$ for both), suggesting that cell-type-specific eRNA can be used to trace lineage-specific TFs.

Given the strong colocalization of enhancer elements with common neuropsychiatric risk variants, we partitioned disease heritability with cell-type-specific eRNAs using linkage disequilibrium (LD) score regression (Finucane et al., 2015). In line with the result from H3K27ac peaks and OCRs, neuron-specific eRNAs were strongly enriched in risk variants for neuropsychiatric traits, including bipolar disorder (BD) (Stahl et al., 2019) and schizophrenia (SCZ) (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2020). Conversely, enrichment in non-neuronal enhancers was not significant. Moreover, we examined the per-SNP heritability of these traits (Methods) and found that the per-SNP heritability of neuronal-specific eRNAs was remarkably higher than that of OCRs and H3K27ac peaks (**Figure 2C** and **2D**), suggesting a role for genetic regulation of cell-type-specific eRNAs in neuropsychiatric disease. Given the relatively small number and genomic coverage of eRNAs (**Figure 2C**), eRNA represents a smaller functional unit that accounts for a higher fraction of disease heritability.

Reproducible dysregulation of eRNA expression in SCZ

To systematically examine the population-level variation of eRNA expression in the human brain, in both healthy and disease states, we leveraged our cell type-specific brain eRNA atlas to quantify eRNAs in a large-scale postmortem brain RNA-seq cohort collected and generated by the CommonMind Consortium (Fromer et al., 2016; Hoffman et al., 2019). A total of 1014 rRNA-depleted total RNA-Seq libraries, covering both the dorsolateral prefrontal cortex (DLPFC, BA9, and BA46) and anterior cingulate cortex (ACC, BA32 and BA24), as well as matched genotyping data from brain banks at the Icahn School of Medicine at Mount Sinai, the University of Pennsylvania and the University of Pittsburgh (a.k.a. MountSinai-Penn-Pitt, the CMC MPP cohort) were used as the discovery set (N_{SCZ} = 254, $N_{Control}$ = 291, **Figure 3A**). We utilized an independent, non-overlapping, cohort consisting of a total of 368 RNA-seq libraries, also derived from DLPFC and ACC, with matching genotype information from the NIMH HBCC brain bank (a.k.a. the CMC HBCC cohort) as a replicate cohort (N_{SCZ} = 78, $N_{Control}$ = 151, **Figure 3A**).

In the discovery cohort, we detected 42,463 eRNAs and 20,704 Ensembl genes expressed at detectable levels (Methods). To identify the dysregulated (DE) transcripts in SCZ, we built a weighted linear mixed model with the normalized combined expression read counts after adjustment for both clinical and experimental covariates (Methods) (Fromer et al., 2016). We found 947 DE genes and 686 DE eRNAs after correction for multiple testing (false discovery rate, FDR ≤ 5%). In the replication analysis, more than 95% of the eRNAs detected in the discovery cohort were replicated (**Figure 3B**, Jaccard index=0.948, $p<10^{-16}$, Fisher's test). Furthermore, the dysregulation of both genes and eRNAs in SCZ are highly consistent between the discovery and replication sets (**Figure 3C**). Overall, our analysis suggests a robust dysregulation of eRNA expression in SCZ.

Given the high reproducibility between the two cohorts, we then performed a joint analysis, leading to the identification of 19,383 genes and 41,984 eRNAs. We controlled for biological factors and technical confounders (**Figure S3A**) and further explored dysregulation in SCZ. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

joint model greatly increased statistical power and, after correction for multiple testing (FDR < 5%), we identified 1,128/932 upregulated eRNAs/genes, and 1,507/1,496 down-regulated eRNAs/genes (**Table S3**). Aside from a substantial percentage of DE eRNAs overlapping with the intronic regions of DE genes, a large fraction of them represent novel loci including both intergenic and non-DE gene intronic regions (**Figure 3D**). We examined the effect of confounders on DE genes and eRNAs with hierarchical clustering (**Figure 3E**). As expected, DE genes and eRNAs exhibited case-control distinctions and were independent of covariates, including RNA integrity number (RIN), post-mortem interval (PMI), age of death (AOD), sex, brain region, and institution. DE genes were also associated with DE eRNAs, suggesting a role for a co-expressed gene-eRNA regulatory network in the etiology of SCZ. In addition, the effect size of DE genes was highly concordant with previously published reports (Gandal et al., 2018; Jaffe et al., 2018), (Gandal et al, SCC= 0.844; Jaffe et al, SCC= 0.516, both $p<10^{-16}$, **Figure S3D**). In line with previous findings, the DEs for both genes and eRNAs had a modest fold change (**Figure S3E**).

We then assessed whether the changes in different types of transcripts reflect the same level of biological information. We estimated a transcriptional perturbation score per individual for each transcript type (**Figure S3F**) and examined concordance among them, as well as the association with case-control status and polygenic risk score for SCZ. The transcriptional perturbation scores are highly consistent between different transcript types and are concordant with diagnoses and polygenic risk scores, indicating that different types of transcripts can capture the same dysregulation information. Interestingly, the genes upregulated in SCZ were enriched for neuron-related cell types and pathways (**Figure 3F**), while the downregulated genes were enriched for various glial (i.e. non-neuronal) cell types and related functions (Lake et al., 2018). In line with gene enrichment, DE eRNAs were also enriched for corresponding single-cell OCR references (Lake et al., 2018) (**Figure 3F**).

Cis-coordination of expression between eRNA and target genes

Given that promoter-enhancer interactions can span considerable genomic distances, and not necessarily in a one-to-one manner, determining the target genes for enhancer elements is still a great challenge (Heinz et al., 2015; Pennacchio et al., 2013; Schoenfelder and Fraser, 2019). Our eRNA expression resource enables direct comparison of expression between enhancers and genes across multiple samples. To link enhancers to target genes, we took into account the joint effect of multiple enhancers, and fit lasso regression models for every gene as a response variable and *cis*-eRNAs within a $\pm$500 Kb window using only control individuals as predictors (**Figure 4A**). The resulting model detected 35,964 gene-eRNA links, consisting of 5,647 genes and 22,147 eRNAs (**Table S4**). On average, a gene was linked to 5 eRNAs (standard deviation, sd=5.34, **Figure S4A**), whereas an eRNA was linked to a single gene (sd=1.09, **Figure S4B**), in agreement with a previous estimation (Delaneau et al., 2019). The majority of the associations are between non-physically-overlapping gene-eRNA pairs (**Figure 4B**). Concordant with the positive regulation of active enhancers, the majority of the eRNA links (75.98%) had positive coefficients. As expected, enhancer-linked genes had a significantly higher expression level relative to the background (expressed genes without linked eRNA) (**Figure S4A**, $p<10^{-16}$, Wilcoxon test). Moreover, eRNAs originating from SEs were more likely to be coordinated with gene expression ($p<10^{-16}$, OR=1.35, Fisher's test); more than 65% of the linked eRNAs come from SE loci,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

highlighting the critical role of SEs in gene regulation. Although we included all eRNAs within a ±500 Kb window to the TSS, the significant eRNAs were located closer to the TSS ($p<10^{-16}$, Wilcoxon test compared to background, **Figure S4C**), even after excluding the physically-overlapping gene-eRNA pairs from the analysis ($p<10^{-16}$, Wilcoxon test compared to background, **Figure S4D**).

Having determined the regulome activity of eRNAs is coordinated with gene expression, we then explored whether changes in the expression of eRNAs alone (within ±500 Kb window from TSS) can predict gene dysregulation in SCZ. We imputed gene DE effect sizes using the eRNA DE effect sizes and the weights from the eRNA-gene lasso regression model. Imputed gene DE effect sizes were highly concordant with the observed gene DE effect sizes (**Figure 4C**, SCC=0.48, $p<10^{-16}$), suggesting that changes in the eRNA activity can explain a large fraction of gene dysregulation in SCZ. In addition, the cell-type specificity between linked pairs was also highly concordant (**Figure 4D**), indicating that our model captures cell-type-specific regulatory mechanisms.

An alternative way of determining promoter-enhancer interactions is by directly measuring the physical contact of DNA using chromosome conformation capture technologies, such as Hi-C. These approaches have found two prominent chromatin structural features that are associated with enhancer-promoter interactions; chromatin loops, which capture long-range interactions; and topological associated domains (TAD), which insulate the interactions across domain boundaries (Yu and Ren, 2017). Compared with background, the eRNA-gene-linked pairs had a significantly higher chance of being linked by chromatin loops (OR=1.84, $p<10^{-16}$, Fisher's test). In addition, most of the identified promoter-enhancer pairs were within the same TAD (59.63%, OR=1.81, $p<10^{-16}$, Fisher's test). We found that the linked eRNA-gene pairs are enriched for being within the same TAD compared to regions an equivalent distance apart (**Figure 4E**). Taken together, our gene-eRNA links are well supported by chromatin interaction features and our approach established cell-type-specific regulatory mechanisms linking eRNAs to their target genes.

Enhancer linked genes are implicated in neuropsychiatric disease

Given that enhancers drive cell-type-specific gene expression, we sought to determine whether eRNA-linked genes are more likely to be cell-type-specific. By examining the *t*-statistics between neuronal and non-neuronal cells, we found that eRNA-linked genes are strongly DE between the two cell types, compared to expressed genes that are not linked ($p<10^{-16}$, Kolmogorov–Smirnov (KS) test, **Figure 4F**). To dissect the function of eRNA-linked genes, we annotated the genes into neuronal and non-neuronal categories based on the DE analysis. As expected, eRNA-linked genes were strongly enriched for the main CNS cell types annotated from single-cell analysis (Lake et al., 2018). The linked neuronal genes were strongly enriched for glutamatergic and GABAergic neurons, and non-neuronal genes were enriched for glial cell types such as oligodendrocytes and astrocytes (**Figure 4G**). In contrast, non-linked genes were not even nominally significant in any of the groups.

Previous studies have highlighted the function of neuronal genes in the etiology of SCZ (Skene et al., 2018). We found, for neuronal genes, that both eRNA-linked and not linked groups were strongly enriched for common neuropsychiatric trait variants, and that the enrichment level was remarkably higher in eRNA-linked groups, especially for SCZ (**Figure 4G**). Aside from the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

neuronal genes, eRNA-linked non-neuronal and non-DE genes were also enriched for the SCZ GWAS signal (**Figure 4G**), indicating that our model captured SCZ functional genes across different cell types. In addition, the neuronal eRNA-linked genes were enriched for SCZ related pathways, including ion transmembrane transport and the Reactome neuronal system (Imbrici et al., 2013); the non-neuronal eRNA-linked genes were enriched for known pathways that are also implicated in SCZ, including ensheathment of neurons and gliogenesis (Dietz et al., 2020). Together, we found that eRNA-linked genes were strongly implicated in SCZ, highlighting the critical role for enhancers in cell-type-specific functions in the CNS and neuropsychiatric disease.

Molecular QTL analysis highlights cell-type-specific enhancer regulation

Given that enhancers outnumber expressed genes, and multiple enhancers could regulate the same gene, eQTL analysis restricted to genes will likely miss critical information from genetic mechanisms mediated by enhancers under certain scenarios. We first examined the expression *cis*-heritability of both genes and eRNAs, which measures the fraction of expression variance explained by SNPs within the *cis*-window. Although lower than that of genes, we observed a substantial proportion of eRNAs are *cis*-heritable ($p<0.05$) (**Figure S5A**). In addition, the heritability of eRNAs was reproducible between the two brain regions (DLPFC and ACC; OR=7.29, $p<10^{-16}$, Fisher's test, **Figure S5C**).

To explore the genetic regulation of enhancers, we performed eQTL analysis for both genes (GeQTL) and eRNAs (EeQTL). For DLPFC and ACC brain regions, gene-eRNA combined expression matrices of European ancestry were independently adjusted for both known and surrogate covariates. An eQTL meta-analysis was subsequently performed with the normalized expression matrices across two brain regions using a linear mixed model to maximize power and account for repeated measures (Zeng et al., 2021). The model identified 3,593,102 *cis*-eQTL variant at FDR ≤ 5%, including 1,001,939 EeQTLs regulating 25,958 (62.86% of autosomal) eRNAs, as well as 2,591,163 GeQTLs regulating 16,165 (86.46 % of autosomal) genes (**Table S5**). The most significant SNP (eSNP) for both genes (GeSNP) and eRNAs (EeSNP) was centered around the corresponding transcription start site (**Figure S5E**). In line with previous findings that QTLs jointly influence multiple molecular phenotypes (Grubert et al., 2015), more than half (56.29%) of the EeQTL encompassed SNPs are also GeQTLs. However, eSNPs were largely independent between genes and eRNAs. Only 6.1% of the EeSNPs reside within any GeSNP LD blocks ($r^2>0.5$) (**Figure 5B**), and only 12.9% of GeSNPs were located in the EeSNP LD blocks ($r^2>0.5$) (**Figure 5C**), highlighting that EeQTLs represent distinct genetic regulation information.

To confirm our eQTL result, we first compared our GeQTL with the GTEx brain tissues (GTEx Consortium, 2020) and observed high π_1 values (median 0.950, **Figure S5F**) as well as the high concordance of allelic effect (**Figure S5F**). We then compared the EeQTL with published brain histone QTL (hQTL) (Sun et al., 2016) and chromatin accessibility QTL (caQTL) (Bryois et al., 2018) data. We observed high π_1 values for EeQTLs (**Figure 5D**). Moreover, both intergenic and intronic EeQTLs exhibited strong positive correlations comparing the effect sizes with caQTL (**Figure 5D**), indicating that EeQTL captures enhancer genetic regulation. We next examined the genetic effect between the linked eRNA and gene pairs. We reasoned that genetic variants would

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

have a concordant effect on functional gene-enhancer pairs. Indeed, we observed the effect sizes were highly consistent between linked genes and eRNAs (**Figure 5E**) and were independent of the overlapped gene-eRNA effect (**Figure S5G**).

Having shown that enhancers play pivotal roles in cell-type-specific gene expression, we next sought to determine whether enhancer loci possess the genetic basis for cell type-specific gene regulation. If this is true, then the cell-type-specific gene eQTL will be enriched in corresponding enhancer regions. To test this, we grouped GeSNPs based on the DE status between neuronal and non-neuronal cells and annotated the SNPs with cell-type-specific chromatin states using GREGOR (Schmidt et al., 2015). As expected, gene eSNPs were strongly enriched for known CREs, including active promoters (TssA) and active enhancers (EnhA) (**Figure 5F**). As distinct from that of the promoters, the enrichment of gene eSNPs at enhancers was highly cell type-specific, highlighting the central role of enhancer elements in cell-type-specific genetic regulation. In contrast to GeSNPs, EeSNPs were strongly enriched in enhancers of corresponding cell types instead of active promoters, distinguishing EeQTLs from GeQTLs.

EeQTL contribute to schizophrenia risk alleles

As described above, eRNA loci are enriched for SCZ GWAS signals (**Figure 2D**). In addition, we found that both EeQTL unique SNPs, and those that are shared with GeQTL SNPs, exhibit an excess of low P values for the SCZ GWAS signal (**Figure 5G**). Taken together, these findings indicate that EeQTLs contribute to SCZ heritability in a complementary manner to GeQTLs. Given the relative independence of eSNP signals, we subsequently performed LD score regression to quantify the GWAS signal enrichment at both EeSNP and GeSNP loci. As expected, we found both EeSNPs and GeSNPs are strongly overrepresented with SCZ common variants as well as other neuropsychiatric traits (**Figure 5H**).

Given that both GeQTLs and EeQTLs are associated with SCZ heritability, we next asked if the heritability is mediated by gene and eRNA eQTL, or if it arises due to other non-causal situations such as linkage and/or pleiotropy. We estimated the proportion of SCZ heritability mediated by gene/eRNA expression ($h_2$med/h_2g) using MESC (Yao et al., 2020). The method quantified the proportion of mediated effect based on the concept that mediation induces a linear relationship between the magnitude of eQTL effect sizes and disease effect size (Yao et al., 2020). We found a substantial fraction (0.068 ± 0.015) of SCZ heritability is mediated by eRNA expression (**Figure 5I**), indicating that eRNA is a key causal factor in the etiology of SCZ. In addition, we combined the gene and eRNA expression matrix and found that the mediated heritability proportion increased from 0.084 ± 0.022 (gene) to 0.13 ± 0.019 (combined), suggesting that genes and eRNAs mediate their genetic effects in an additive manner.

Collectively, our analysis highlights the critical and independent role played by EeQTLs in neuropsychiatric disease, comparable to that of GeQTLs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Adding eRNAs to transcriptome-wide association studies facilitates fine-mapping and interpretation of SCZ GWAS loci

Having shown that SCZ genetic variants contribute substantially to gene and eRNA expression, we next sought to identify the potential genes and eRNAs that mediate the genetic risk for SCZ. We performed transcriptome-wide association studies (TWAS) (Gusev et al., 2016) using gene and eRNA eQTLs, and the most recent SCZ GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2020). Initially, we built elastic net and lasso regression genetic variant-based expression prediction models (Barbeira et al., 2018) for *cis*-heritable transcripts in DLPFC and ACC brain regions. This resulted in expression models for 10,669 unique genes and 8,702 unique eRNAs from the two brain regions, which markedly increased transcriptome coverage for TWAS compared to previous human brain studies in SCZ (Gandal et al., 2018; Gusev et al., 2018; Huckins et al., 2019; Zhang et al., 2019). The models yielded 204 genes and 98 eRNAs that are outside of the MHC region and are associated with SCZ ($P_{bonferroni}$<0.05), covering 104 of 264 non-MHC autosomal independent genome-wide significant SCZ GWAS loci (**Figure 6A**). Specifically, 26 loci are shared between eRNAs and genes, 23 loci are only tagged by eRNAs and for 55 loci only genes were detected. To confirm the reproducibility of our studies, we compared the TWAS *Z* scores between the two brain regions, as well as the *Z* scores between our gene models and previous reports (Gandal et al., 2018; Zhang et al., 2019), where both comparisons exhibit high concordance (**Figure S6A** and **S6B**).

To avoid spurious associations from LD structure (Wainberg et al., 2019), we first checked if the eQTL and GWAS association is driven by the same causal variants. We conducted a colocalization analysis (Giambartolomei et al., 2014, 2018) to estimate the probability that the eQTL and GWAS signals are associated (PP4) or not (PP0, PP1, PP2 and, PP3). We found both genes and eRNAs exhibited high PP4 value (**Figure 6B**), 66 of the 98 TWAS eRNAs, and 139 of the 202 TWAS genes colocalized between eQTL and GWAS signals (PP4>0.5). We next conducted a fine-mapping analysis that controls for the correlation structure introduced by LD and SNP weights, as well as certain pleiotropic effects to refine the TWAS associations using FOCUS (Mancuso et al., 2019). To account for genes that are filtered by low *cis*-heritability, we included gene prediction results from GTEx expression panels. The analysis yielded a credible set of 466 transcripts, including 384 genes and 82 eRNAs covering 151 GWAS loci, where 20 of the loci were only tagged by eRNAs and 29 of the loci were tagged by both gene and eRNAs.

To further refine the transcript-based fine-mapping of the TWAS outcome, we required the genes and eRNAs to (i) have $P_{bonferroni}$ < 0.05 in the TWAS analysis, (ii) have PP4 > 0.8 in the colocalization analysis, and (iii) be within the credible sets of FOCUS transcripts, resulting in 16 eRNAs and 56 genes (**Table S6**). Although a substantial percentage of the associated genes have been identified previously, many of them are novel (**Figure 6C**). Using our gene-eRNA link model, we identified 19 genes (MAN2A1, VRK2, FANCL, LINC01877, FTCDNL1, DPYD, CTNND1, MFGE8, ABHD2, STH, SATB2, C2orf69, MAIP1, PJA2, FAM114A2, MMP16, KIF7, AC092691.1, SAP30L) that are associated with the fine mapped eRNAs (**Figure 6C** and **Table S6**), many of which have been reported previously (**Figure 6C**) but have not been directly identified by TWAS analysis (Gandal et al., 2018; Wang et al., 2018a, 2019).

We provide an illustrative example for an SCZ locus that includes eRNA41216 (**Figure 6D**), a novel eRNA that resides within a super-enhancer. The GWAS locus is associated with 12 eRNAs

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and 32 genes but only eRNA41216 is identified as a causal gene by fine-mapping. The top SNP of this eRNA, rs2247233, is also one of the top GWAS SNP (**Figure S6C**). In addition, we performed a conditional analysis and found that the eRNA41216 TWAS signal fully explained the GWAS significance (**Figure 6D**, bottom).

The eRNA is linked to three genes, ABHD2, MFGE8, and KIF7. To experimentally validate the enhancer target, we performed CRISPR interference of eRNA41216 in neural progenitor cells (**Figure 6D**). This led to a 20-30% reduction in gene expression for KIF7 and ABHD2, two genes that reside ~200kb from the eRNA, whereas levels of MFGE8, 450kb from the eRNA, remained unchanged. KIF7 has been known to play a role in regulating neuronal development and brain abnormalities in ciliopathies (Liem et al., 2009), but was not significant in the TWAS analysis (TWAS $P_{bonferroni}$>0.1, FOCUS posterior inclusion probability, PIP<0.001). Moreover, knockdown of KIF7 results in thinner and shorter processes of multipolar neurons (Guo et al., 2015). ABHD2, which is not significant in the TWAS analysis (TWAS P>0.1, FOCUS PIP<0.001), encodes an enzyme that catalyzes the hydrolysis of 2-arachidonoyl glycerol, which is a signaling lipid in the CNS and a key regulator of neurotransmitter release (Baggelaar et al., 2018; Ogasawara et al., 2016).

Trans-coordination: gene-eRNA co-expression network dissects SCZ etiology

Higher-order chromatin interactions and coordinated regulation of gene expression are critical to brain development and function. The organizational principles of eRNA-gene transcriptional regulation can be captured based on network analysis. One expectation of polygenic inheritance of the disease is that the coordination of *cis*-linked genes (N=5,647) and eRNAs (N=22,147) may be subtly altered in individuals with SCZ. To assess this, we constructed a co-expression network using the multiscale embedded gene coexpression network analysis (MEGENA) (Song and Zhang, 2015). MEGENA recovers a network that consists of nodes (genes and eRNAs) and edges connecting nodes and assigns genes and eRNAs into subnetworks called modules.

A total of 153 co-expressed modules were identified, the majority of which included both genes and eRNAs (**Figure 7A**). To prioritize SCZ implicated modules, we assessed the enrichment with SCZ DE signatures and SCZ GWAS loci (Methods). We identified 9 SCZ up-regulated modules, 5 SCZ down-regulated modules, and an SCZ GWAS enriched module, M26 (**Table S7**). Interestingly, eRNA41216, *KIF7*, and *ABHD2*, which were prioritized based on TWAS, are all within the SCZ GWAS associated module M26. We subsequently examined the biological function of these 15 prioritized modules. Given that the modules consisted of a substantial percentage of eRNAs, for every module, we expanded the gene list to also include the eRNA-linked genes. As expected, the modules were strongly enriched for CNS cell types and related biological processes (**Figure 7B**). In line with previous findings (Gandal et al., 2018), modules that were up-regulated in SCZ were enriched for glutamatergic and GABAergic neurons, as well as neuronal and cognition-related biological processes. Of the five down-regulated SCZ modules, four were enriched for oligodendrocyte cells and related functions, such as cytoskeleton organization and ensheathment of neurons. The remaining down-regulated SCZ module, M103, was enriched for microglia and was associated with immune regulation. We observed a similar, but less significant, enrichment after excluding eRNA-linked genes (**Figure S7B**). We then

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

examined the overlap of intramodular eRNAs with adult and fetal enhancers (Li et al., 2018). Three out of the nine modules that were up-regulated in SCZ were overrepresented for fetal enhancers, while all five down-regulated modules were enriched for adult-specific enhancers. In line with the enrichment of eRNAs with fetal enhancers, we also observed that corresponding genes (intramodular genes and eRNA-linked genes) exhibited distinct expression trajectories during fetal development (**Figure S7C**).

The evolutionary expansion of the cerebral cortex contributes to higher cognition in humans. We next explored the evolutionary conservation of brain eRNAs. We examined the eRNA PhyloP conservation scores as well as the human gained enhancers (HGE) (Reilly et al., 2015) (Methods). Interestingly, module M26, which was overrepresented for SCZ GWAS signal, was strongly enriched for low-PhyloP-eRNAs and HGEs. In one example, eRNA40115, a novel eRNA, is predicted to be a driver of M26. It exhibits an extremely low PhyloP score ($1.02x10^{-03}$) and overlaps with HGE. Based on our eRNA-gene link models, eRNA40115 regulates three genes; GABRA5, GABRB3, and GABRG3, which encode subunits of the GABA-A receptor, known to play a pivotal role in neuropsychiatric disease (Rudolph and Möhler, 2014).

Lastly, we set out to determine the potential driver TFs that regulate the co-expression modules (**Figure 7C**). We examined TF motif enrichment as well as the correlation between the TF gene and module eigengenes. Interestingly, most of the resulting TFs have been suggested to play roles in neuronal function and psychiatric disease. A salient example is the TF driver of the SCZ GWAS enriched module, M26, JUN, which is a part of AP-1 proteins, and has long been known to play a critical role in CNS development and is implicated in neuropsychiatric disease (Mahar and Cavalli, 2018; Raivich and Behrens, 2006).

Discussion

A growing body of evidence suggests that enhancer elements play a pivotal role in neuropsychiatric diseases (Finucane et al., 2015; Sullivan and Geschwind, 2019; Wang et al., 2018a). However, a systematic view, including a cell-type-specific reference map of eRNAs, dysregulation in disease, regulatory circuits, and genetic effects of enhancer sequences is lacking. Here, we utilized population-level variation of eRNA and gene expression in the human brain to provide a comprehensive assessment of the regulatory mechanisms of transcribed enhancers in SCZ. We illustrate how to leverage large-scale transcriptome data to investigate enhancer regulatory circuits and provide novel insights into complex traits.

Given that eRNA is modestly expressed, previous attempts to identify eRNAs using a single assay either used a stringent threshold and missed the majority of eRNAs (Andersson et al., 2014), or, with a more flexible cutoff, identified mostly non-enhancer elements (Dong et al., 2018). In our analysis, we used a step-wise approach to define brain eRNAs with increased sensitivity and specificity. We initially performed cell-type-specific eRNA mapping based on supervised machine-learning integrative analysis of multi-omics data. We then further refined eRNA mapping and studied the properties of the eRNA regulome using population-level expression from RNA-seq data. This approach is very sensitive, resulting in 30,795 neuronal and 23,265 non-neuronal eRNAs, expanding, by an order of magnitude, the repertoire of known transcribed enhancers in the human brain. In terms of specificity, the accuracy of our method is confirmed by active

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

enhancer histone modification occupancy (~95% H3K27ac), bidirectional transcription initiation signal, as well as the high precision-recall of our predictive model. Compared to the traditional and broad active enhancer assays, H3K27ac and H3K4me1 ChIP-seq, our eRNA-based approach has the advantage of higher resolution to define smaller functional regions as well as the underlying TF binding sites. In addition, eRNA can better refine the chromatin accessibility landscape by subsetting ATAC-seq peaks that represent active enhancers from other regulatory sequences. In line with this, analysis of cell type-specific eRNAs identifies markedly higher SNP heritability for neuropsychiatric traits than regulatory sequences defined using ATAC-seq and H3K27ac ChIP-seq.

Based on the coordinated expression of eRNAs and target genes, we utilized population-level variation of the transcriptome and regulome in 1,382 brain samples to model *cis*-co-expression and determine gene-eRNA interactions. As enhancers can exert their effects across long genomic distances, the closest gene is not necessarily the target gene. Spatial interactions determined by Hi-C are limited by the resolution of the method and by bias due to proximity ligations (Fudenberg and Imakaev, 2017), and may have other functions besides facilitating promoter-enhancer interactions. Our approach is validated by successfully imputing gene dysregulation with eRNA information, concordance with chromatin organization features, consistent allelic genetic effects, and *in vitro* CRISPR inference validation. eRNA-linked genes are highly expressed, are strongly enriched for CNS cell types and neuropsychiatric trait risk loci, highlighting the importance of eRNAs in the etiology of disease.

We found highly reproducible changes in the expression levels of eRNAs and genes in cases with SCZ compared to controls. These changes affected a large number of transcripts and were subtle, which is consistent with the polygenic nature of the disease (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2020) and the transcriptome dysregulation underlying SCZ (Fromer et al., 2016). We performed network analysis of transcriptome and regulome data to create modules that capture the organizational principles of eRNA-gene interactions (Song and Zhang, 2015). To functionally annotate each module, we evaluated the cell type, biological process, development stage, evolutionary conservation, and enrichment with SCZ transcriptional perturbations and genetic risk variation. We identified robust changes affecting *cis*- and *trans*-eRNA-gene co-regulation in SCZ. Consistent with previous findings (Gandal et al., 2018; Roussos et al., 2012), these changes included modules that showed a preference for cell type-specific markers and fetal enhancers. We highlighted a neuronal module, M26, with high enrichment for SCZ genetic signal and human evolved enhancers. eRNA40115, a novel eRNA, represents a human evolved enhancer and functions as a key driver for M26, regulating three genes (*GABRA5*, *GABRB3,* and *GABRG3*) that encode subunits of the GABA-A receptor. The GABA-A receptor is a pentameric heterodimer composed of a combination of 19 subunits (Wafford et al., 2004). The α5-containing GABA-A receptor represents less than 5 percent of all GABA-A receptors and has a restricted distribution within the prefrontal cortex, hippocampus, and amygdala (Gill and Grace, 2014). Abnormalities in the GABA-A receptor have been observed extensively in SCZ (Marques et al., 2020; Rudolph and Möhler, 2014) and allosteric modulation of the α5-containing GABA-A receptor has been proposed to have the potential to ameliorate cognitive deficits in SCZ (Gill and Grace, 2014).

Given the significant *cis*-heritability of eRNA expression, we performed eQTL analysis to link *cis* genetic variation to eRNAs (Zeng et al., 2021). We found that EeQTLs were (i) highly

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

concordant hQTLs and caQTLs, (ii) centered around the eRNA loci, (iii) relatively independent of GeQTLs, and (iv) enriched in active enhancers, suggesting that eRNAs represent enhancer activity instead of transcriptional noise. Both eRNA loci and EeSNPs confer substantial neuropsychiatric disease liability. We subsequently determined the proportion of SCZ heritability mediated in *cis* by eRNA expression (Yao et al., 2020) and found that eRNA expression alone, or combined with gene expression, explained a substantial percentage of SCZ *cis*-genetic variants (6.8% and 13%, respectively). We performed TWAS and fine-mapping and found that a significant fraction of the SCZ GWAS loci is captured only by eRNAs. eRNA41216, another eRNA within the M26 module, was prioritized based on fine-mapping in an SCZ locus containing 12 eRNAs and 32 genes. eRNA41216 was predicted to regulate three genes, among which two (*KIF7* and *ABHD2*), located ~200kb from the eRNA, were validated using CRISPR interference in neural progenitor cells. Taken together, our study further emphasizes the importance of the regulome as an additional layer to functionally characterize disease vulnerability.

Overall, the consideration of population variation in eRNA expression in large-scale total RNA-seq analysis provides novel insights into the regulatory mechanisms of gene expression, as well as the genetic effects influencing complex traits. The framework described in this study provides an inexpensive and high-resolution approach to explore active enhancer function in most human tissue and cell types.

CONSORTIA

The members of the CommonMind Consortium who are not listed in the primary author list are Andrew Chess, Attila Gulyás-Kovács, Bibi Kassim, Eva Xia, Joseph D Buxbaum, Laura Sloofman, Lizette Couto, Mariana Amaro, Marina Iskhakova, Michael Breen, Olivia Devillers, Schahram Akbarian, Shan Jiang, Steven P Kleopoulos, Yixian Ma, Yungil Kim, Sabina Berretta, Ajeet Mandal, Barbara K Lipska, Francis McMahon, Pavan K. Auluck, Stefano Marenco, Kelsey S Montgomery, Mette A Peters, Solveig K Sieberts, Chang-Gyu Hahn, Raquel Gur, Jiebiao Wang, Bernie Devlin, David A Lewis, Lambertus Klei, Enrico Domenici, Michele Filosi, Roberto Visintainer, Douglas M Ruderfer, and Lide Han.

ACKNOWLEDGMENTS

We thank Dr. Won Min Song for valuable discussions about MEGENA network analysis. We thank the computational resources and staff expertise provided by the Scientific Computing of the Icahn School of Medicine at Mount Sinai.

The CommonMind data sets were generated as part of the CommonMind Consortium supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH- 075916, P50M096891, P50MH084053S1, R37MH057881, AG02219, AG05138, MH06692, R01MH110921, R01MH109677, R01MH109897, U01MH103392, U01MH116442, project ZIC MH002903 and contract HHSN271201300031C through IRP NIMH. Brain tissue for the study was obtained from the following brain bank collections: The Mount Sinai/JJ Peters VA Medical Center NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh Brain Tissue Donation Program, and the NIMH

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Human Brain Collection Core. CMC Leadership: Panos Roussos, Joseph Buxbaum, Andrew Chess, Schahram Akbarian, Vahram Haroutunian (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur (University of Pennsylvania), Chang- Gyu Hahn (Thomas Jefferson University), Enrico Domenici (University of Trento), Mette A. Peters, Solveig Sieberts (Sage Bionetworks), Stefano Marenco, Barbara K. Lipska, Francis J. McMahon (NIMH).

This work is supported by the National Institute on Aging, NIH grants R01-AG067025 (to P.R. and V.H.), R01-AG065582 (to P.R. and V.H.) and R01-AG050986 (to P.R.). Supported by the National Institute of Mental Health, NIH grants, R01-MH110921 (to P.R.), U01-MH116442 (to P.R. and V.H.), R01-MH125246 (to P.R.), R01-MH106056 (to P.R. and K.J.B.), R01-MH109897 (to P.R. and K.J.B.) and R01-MH121074 (to K.J.B.). Supported by the Veterans Affairs Merit grant BX002395 (to P.R.). P.D. was supported in part by NARSAD Young Investigator Grant 29683 from the Brain & Behavior Research Foundation. G.E.H. was supported in part by NARSAD Young Investigator Grant 26313 from the Brain & Behavior Research Foundation. J.B. was supported in part by NARSAD Young Investigator Grant 27209 from the Brain & Behavior Research Foundation.

AUTHOR CONTRIBUTIONS

P.R. conceived of and designed the project. J.F.F. and P.R. designed experimental strategies for epigenome profiling of human postmortem tissue. J.F.F. prepared nuclei and performed FANS. J.F.F. and R.M. generated ATAC-seq data. P.A. generated the ChIP-seq and Multi-omics RNA-seq data. S.R. generated Hi-C data. S.R., J.M.V., M.B.F., K.G.T. and K.J.B. performed the CRISPR interference experiments. P.D. and P.R. designed analytical strategies. J.B. K.G. and P.D. conducted initial bioinformatics, sample processing and quality control for the Multi-omics cohort. G.E.H. and W.Z. conducted initial bioinformatics, sample processing and quality control for the CMC cohort. P.D. developed the computational scheme and performed the downstream analysis. B.Z. performed the MMQTL analysis. G.V. performed the polygenic risk score analysis. V.H. provided human brain tissue. P.D. and P.R. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Main figures

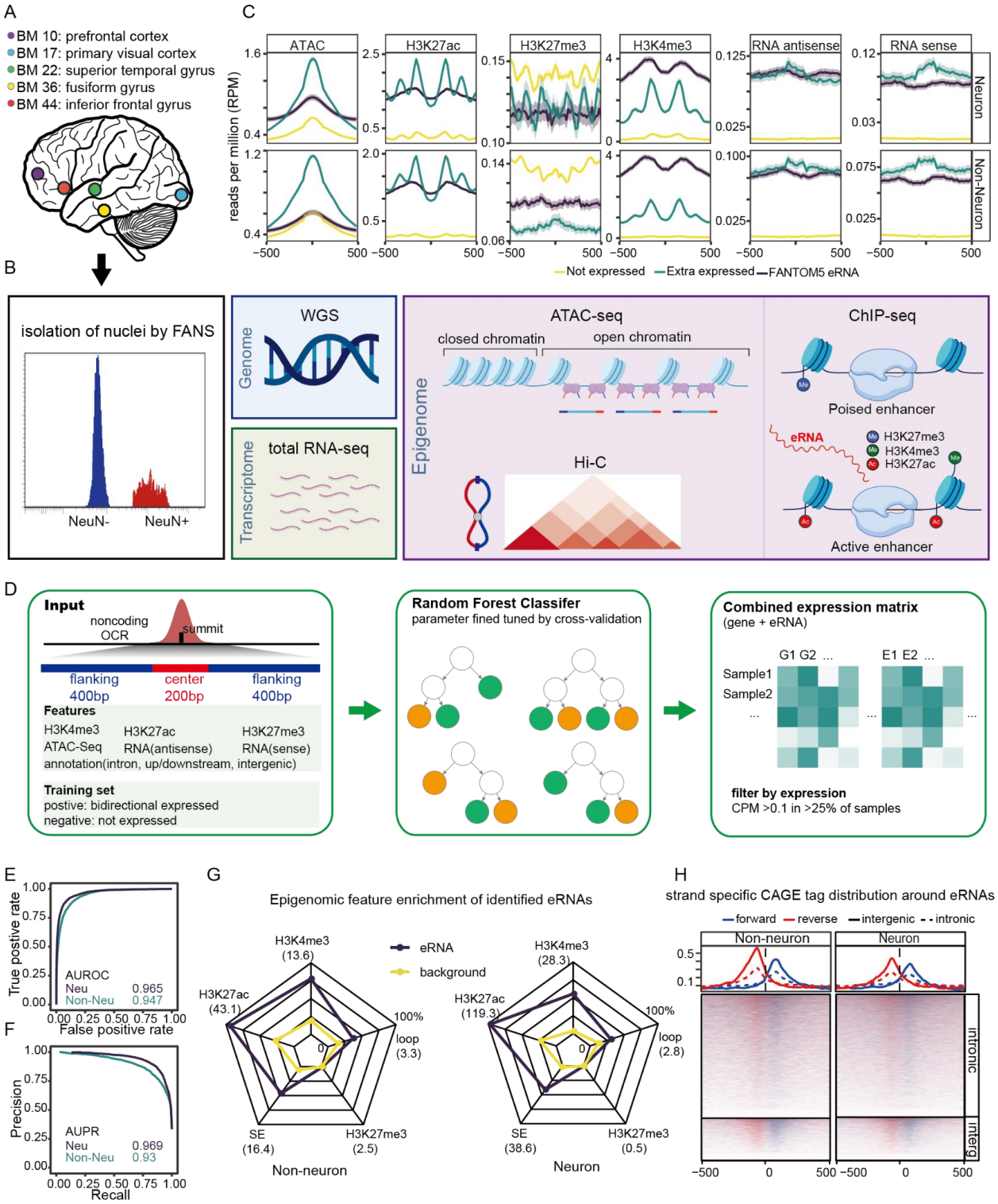


Figure 1. Catalog cell-type-specific enhancers in the human cortex

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A) Dissections from five brain regions (Brodmann area 10, 17, 22, 36, and 44) of 10 control subjects were obtained from frozen human postmortem tissue.
B) Combined with FANS, we performed functional assays including total RNA-seq, ATAC-seq, H3K4me3/H3K27ac/H3K27me3 ChIP-seq, and Hi-C for neuronal (NeuN+) and non-neuronal (NeuN-) nuclei.
C) Transcriptomic and Epigenomic profiles around expressed enhancers (FEs and EEs), and not expressed enhancers (NEs), the shadow shows the 95% confidence intervals.
D) Flow charts demonstrate eRNA identification pipelines. Briefly, transcriptomic and epigenomic signals around the enhancer regions were aggregated for both neuronal and non-neuronal cells. Expressed enhancers (FEs and EEs) were used as positive sets, and not expressed enhancers(NEs) were used as negative sets. The parameter was tuned with 10-fold cross-validation to select a random forest model to classify expressed and not expressed enhancers. Lastly, the enhancer expression matrix and gene expression matrix were combined for downstream analysis.
E) ROC (true positive rate vs false positive rate) and
F) PR (precision-recall) curve of the resulting random forest models exhibits high AUROC (area under the precision-recall curves) and AUPRC (area under the precision-recall curves).
G) The radar plots show typical enhancer-related signals including H3K4me3, H3K27ac, H3K27me3, super-enhancer, and loop anchor occupancy between identified eRNAs and background. The value within parentheses indicates the odds ratio between identified eRNA and background (not-expressed enhancers).
H) Strand-specific CAGE tag average profile (top) and distribution (bottom) for intergenic and intronic eRNAs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

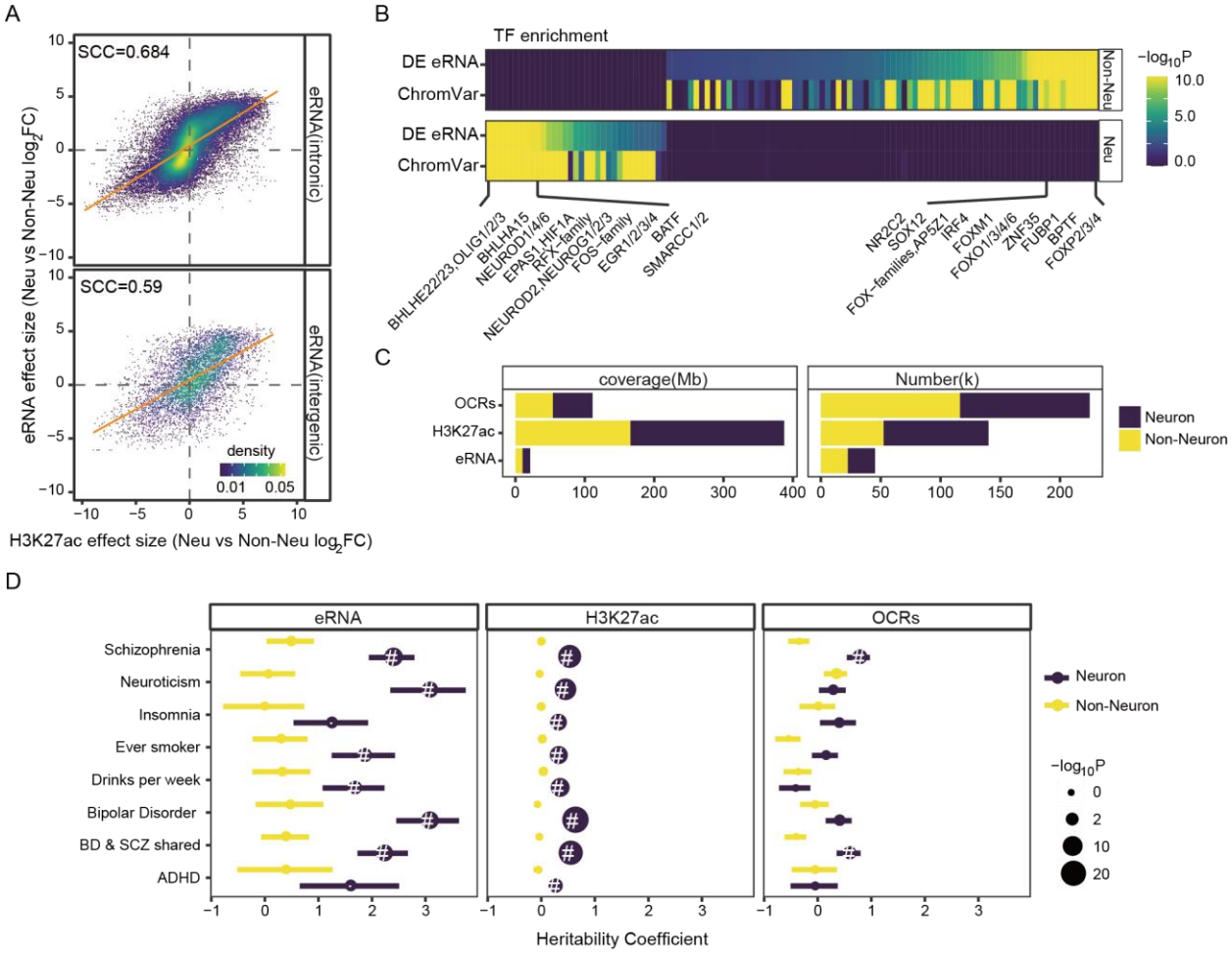


Figure 2. eRNA expression captures cell-type-specific enhancer function

A) cell-type-specific effect size ($\log_2$ fold change) between eRNAs and overlapping H3K27ac peaks are highly consistent for both intergenic and intronic eRNAs. SCC represents the Spearman correlation coefficient (ρ) ($p<10^{-16}$ for both).

B) The significant TF motifs that are determined by DE eRNAs. Heatmap shows the $-\log_{10}P$ values from DE eRNA TF enrichment and ATAC-Seq ChromVar analysis. Top 10 enriched TFs are labeled.

C) DE eRNA/peaks coverage (Mb) and numbers (k)

D) LD score regression enrichment for DE eRNA/peaks of different neuropsychiatric traits. A positive coefficient signifies enrichment in heritability (per base enrichment). "·": Nominally significant ($p<0.05$); "#": significant after FDR (Benjamini & Hochberg) correction ($FDR<0.05$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

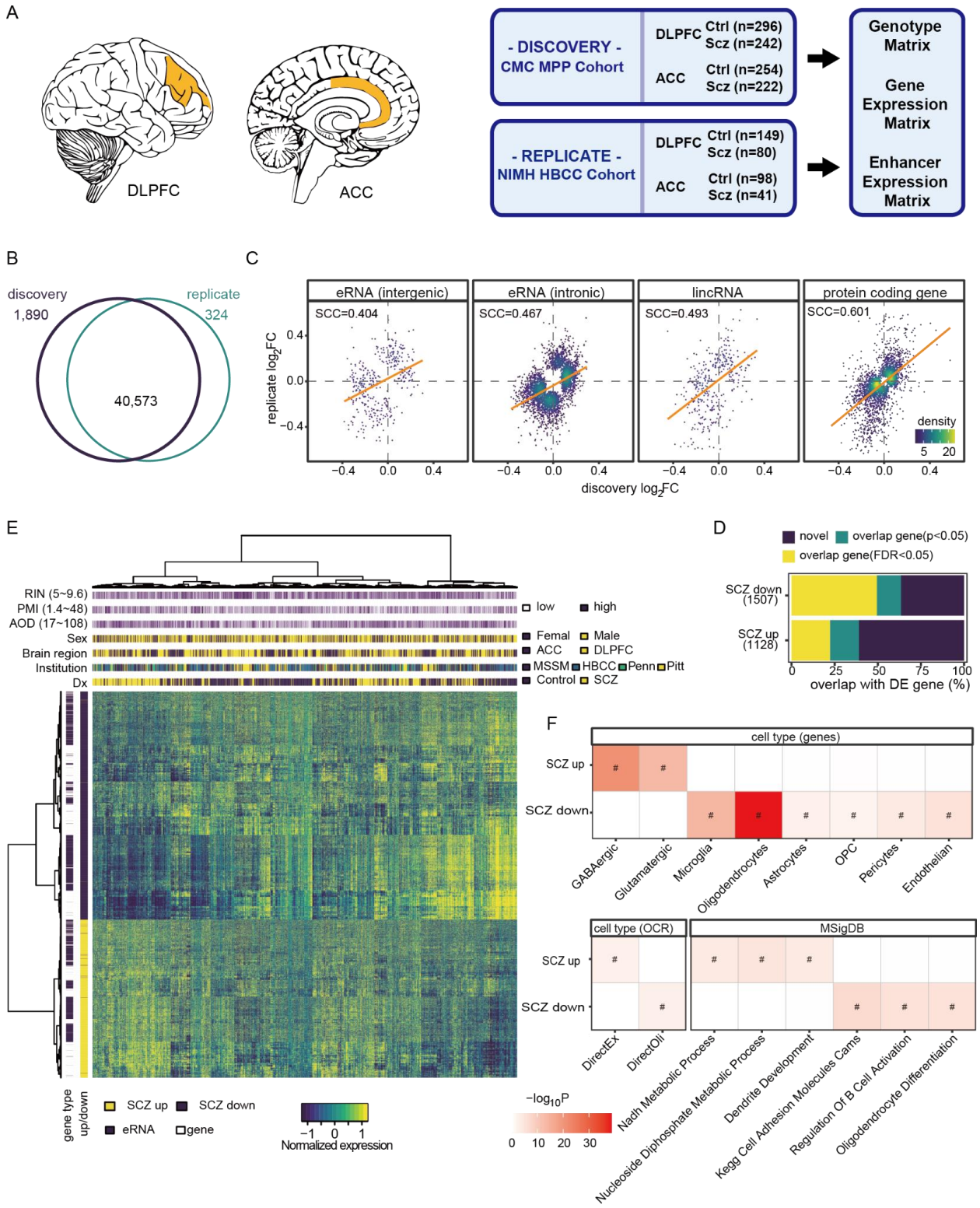


Figure 3. Reproducible dysregulation of eRNA expression in SCZ

A) Two independent Cohorts were used for this study, both cohorts consisting of total RNA-seq from DLPFC and ACC postmortem brains as well as matched genotyping data. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

discovery set includes 1,014 RNA-seq samples from N_{SCZ}=254 and $N_{control}$=291 control unique individuals. The replicate set includes 368 RNA-seq samples from N_{SCZ}=78 and $N_{control}$=151 unique individuals. Brain regions were visualized with cerebroViz (Bahl et al., 2017).

B) The Venn diagram shows the overlap of detected eRNAs between the discovery and the replicate analysis.

C) Effect sizes ($\log_2$ Fold change) between nominal significant transcripts ($p<0.01$ in at least one set) are highly consistent between the discovery and the replicate set. SCC represents the Spearman correlation coefficient (ρ)(all $p<10^{-16}$).

D) Percentage of DE eRNAs overlapped with DE genes ($p<0.05$ and FDR<0.05).

E) Bivariate clustering of individuals (columns) and gene/eRNAs (rows) depicting the case vs. control differences of the 2,428 DE genes and 2,635 DE eRNAs. Bars on the top indicate the Dx (diagnosis), brain bank Institution, brain region, age of death (AOD), PMI, and RIN. Bars on the left indicates transcript type, and effect direction (up or down dysregulated)

F) Overlap between dysregulated genes and cell type marker genes (Lake et al., 2018)(top), between dysregulated genes and GO terms (Liberzon et al., 2011) (bottom right), and between dysregulated eRNAs and cell type marker OCRs (Lake et al., 2018) (bottom left). "#": one-sided binomial FDR < 0.05.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

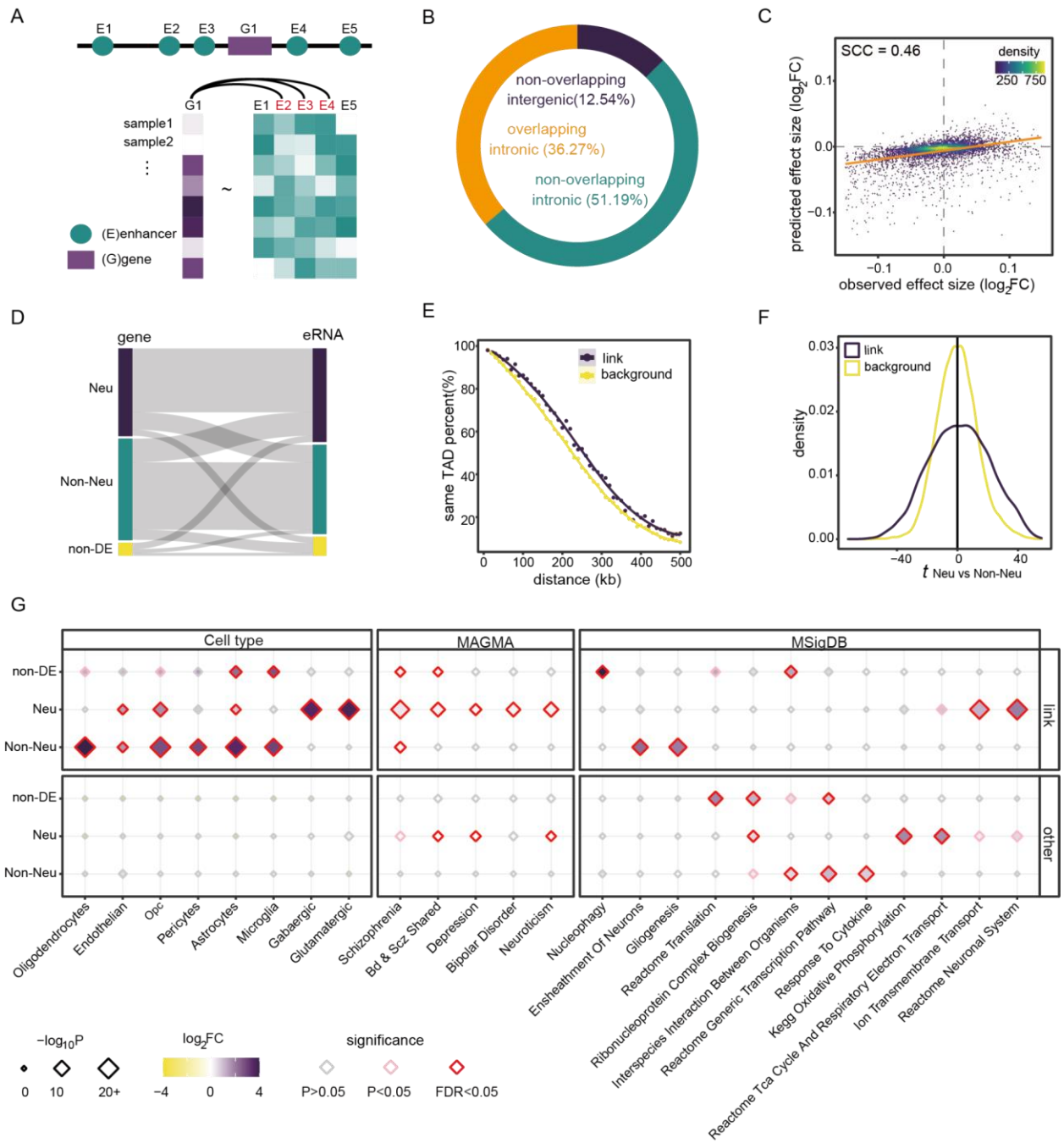


Figure 4. eRNA regulation of target genes

A) Demonstration of the lasso link model. For every gene, we considered all the eRNAs within ±500kb of the TSS and fit a lasso model to select the linked eRNAs.

B) Distribution of the three different classes of eRNA-gene pairs: non-overlapping intergenic-eRNA gene pairs, overlapping intronic-eRNA gene pairs, and non-overlapping intronic-eRNA gene pairs

C) Scatter plot of SCZ-control effect size between observed in gene dysregulation and imputed from linked eRNA effect sizes. SCC represents the Spearman correlation coefficient (ρ) ($p<10^{-16}$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

D) The differential expression status in the two cell types between gene and linked eRNAs.

E) Percentage of gene promoter and eRNA within the same TAD for different genomic distances at 10kb intervals.

F) Compared to the background, the eRNA-linked genes have remarkable higher absolute *t* statistics between neuronal and non-neuronal cells (KS test, $p<10^{-16}$)

G) Cell types, neuropsychiatric common variants, and biological pathways enriched at different classes of linked and not-linked genes.

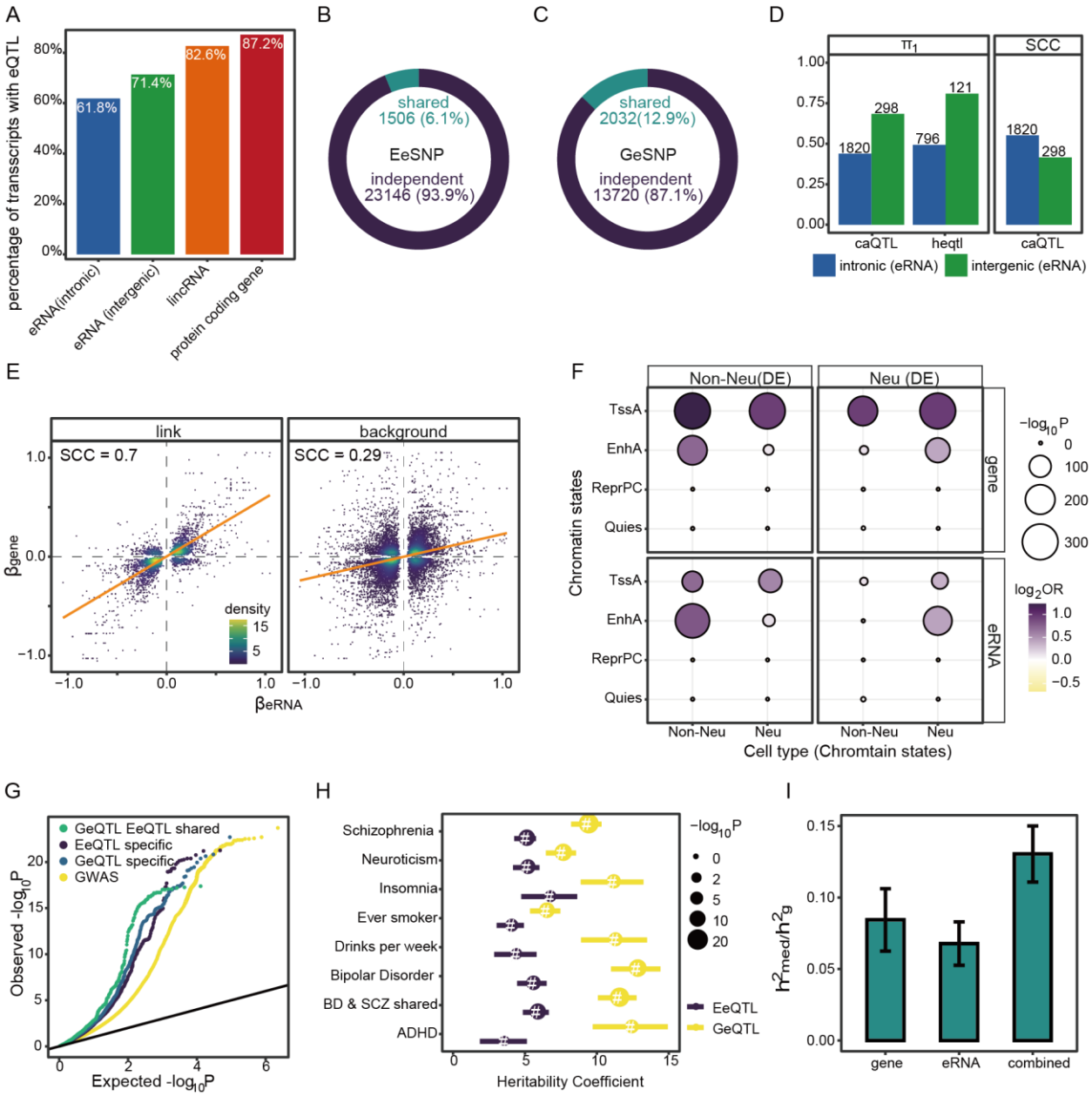


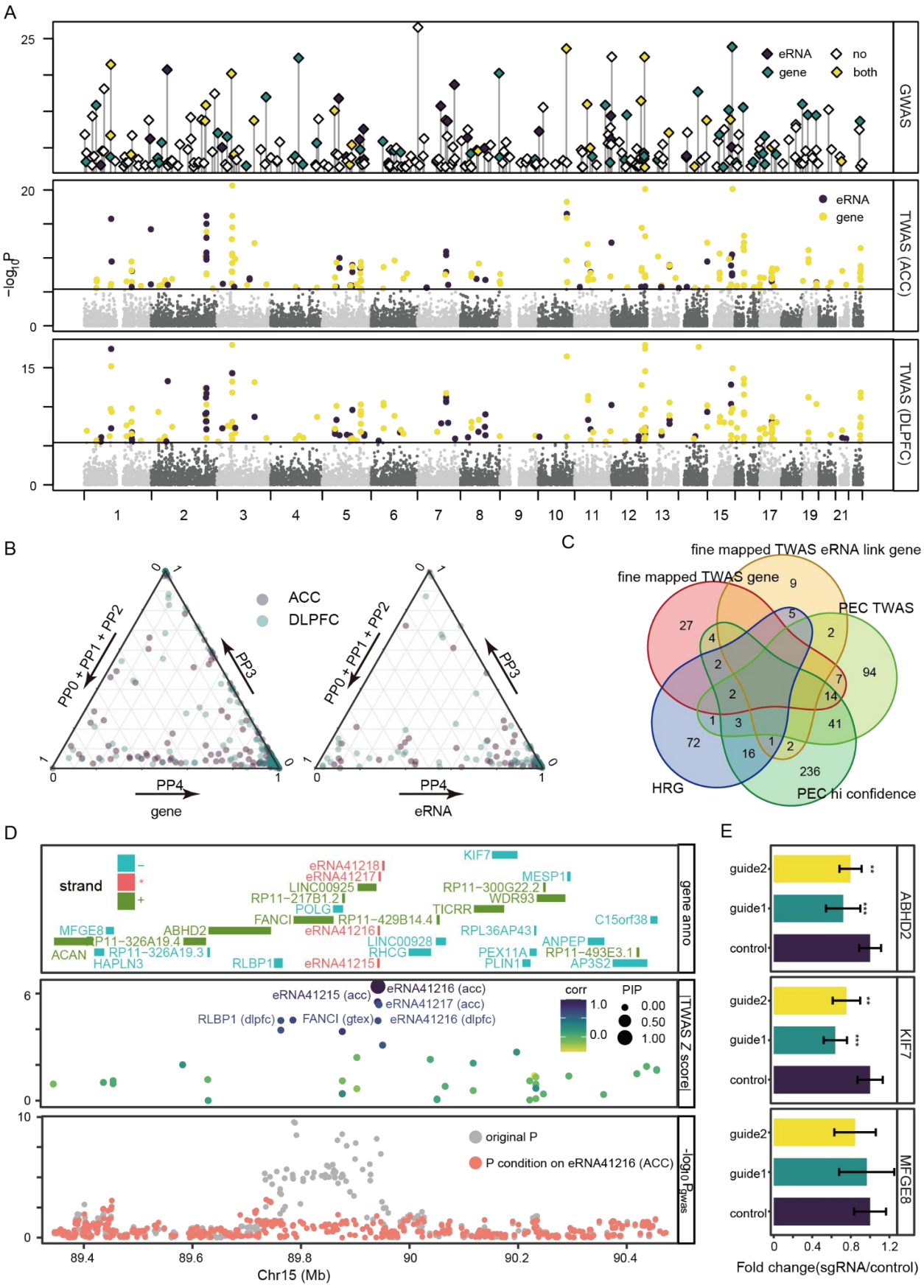
Figure 5. Genetic effects of eRNAs

A) The percentage of different classes of autosome transcripts that have significant eQTLs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

B) Percentage of EeSNP reside within any GeSNP LD blocks (r^2>0.5)
C) Percentage of GeSNP reside within any EeSNP LD blocks (r^2>0.5)
D) The replication of reported hQTL and caQTL in our analysis. Storey's π_1 values for significant hQTL and caQTL in the EeQTLs. SCC values are the Spearman correlations of effect sizes between CaQTL and corresponding EeQTL. The size of the point corresponds to the number of unique eRNAs used. The number above the bar corresponds to the count of unique eRNAs used.
E) The allelic genetic effects between gene and target eRNAs are highly consistent. SCC represents the Spearman correlation coefficient (ρ)
F) Enrichment of neuronal and non-neuronal genes and eRNAs corresponding eQTLs in four chromatin states: TssA (active promoter), EnhA (active enhancer), ReprPC (polycomb repression), and Quies (other). The filled color represents enrichment fold change, and the size corresponds to enrichment P values.
G) Quantile–quantile plot of SCZ GWAS p values. EeQTL specific, eQTL specific, and shared SNP are shown in comparison with genome-wide SNPs. GWAS SNPs were binarily annotated using SNPs within r^2 >0.8 of the eSNP.
H) LD score regression enrichment for gene and eRNA eSNPs of different neuropsychiatric traits. Positive coefficient signifies enrichment in heritability (per base enrichment).”·”: Nominally significant (p<0.05); ”#”: significant after FDR (Benjamini & Hochberg) correction (FDR<0.05)
I) Estimated proportion (± standard error) of heritability mediated by the *cis* genetic component of assayed eRNA, gene, and combined expression for SCZ.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 6. SCZ TWAS

A) Manhattan plot of SCZ TWAS enrichment in DLPFC, ACC, and the independent genome-wide significant SCZ associations (excluding chrX and MHC). GWAS node height corresponds to the index SNP significance, and the color indicates if the GWAS loci are associated with eRNA, gene, both gene, and eRNA, or none of them. Significant TWAS eRNA (purple) and genes (yellow) are highlighted in different colors.

B) Ternary plots showing coloc posterior probabilities for significant TWAS genes and eRNAs respectively. PP0+PP1+PP2: three scenarios for lack of test power; PP3: independent causal variants; PP4: colocalized causal variants.

C) Venn diagrams show the overlap between the fine mapped TWAS genes, TWAS-eRNA linked genes, PsychENCODE (PEC) TWAS genes (Gandal et al., 2018), PsychENCODE(PEC) high confidence risk genes (Wang et al., 2018a), and SCZ high-confidence risk genes (HRG) (Wang et al., 2019).

D) Illustration of genomic loci at chr15, harboring multiple TWAS loci. Top panel, transcript position (only *cis*-heritable transcripts are shown), the color indicates transcription direction of genes. Middle, gene/eRNA TWAS *Z* score absolute value, point size indicates the FOCUS posterior inclusion probability (PIP), color indicates the genetic correlation with the highest PIP. Bottom, Manhattan plot of SCZ GWAS signal, before and after conditioning on eRNA41216.

E) KRAB-dCas9-mediated repression of eRNA41216 leads to the reduction of ABHD2 and KIF7 expression as measured by qPCR in neural progenitor cells. (*** $p<0.001$, ** $p< 0.01$, ANOVA, N=6, mean ± standard error are shown)

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

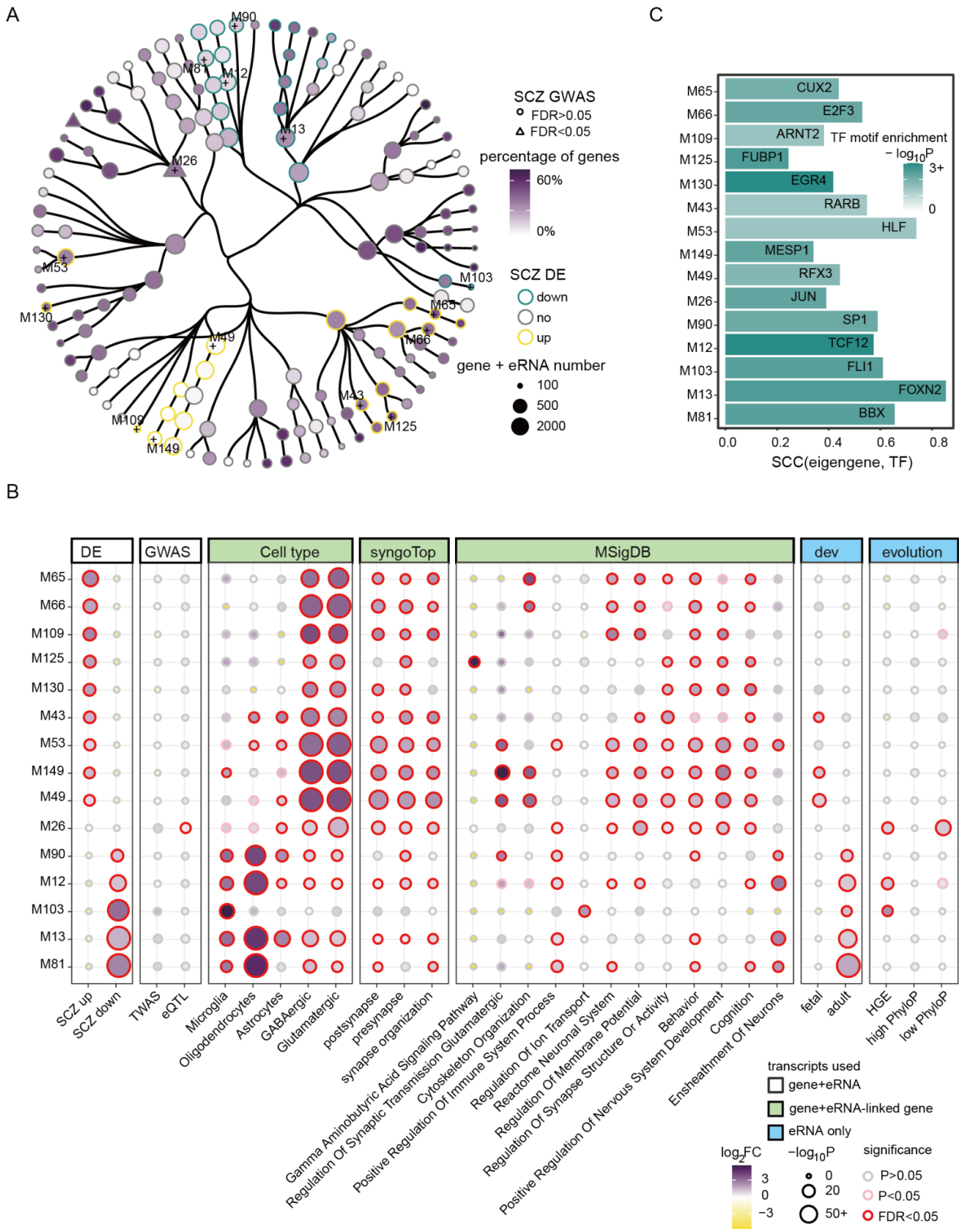


Figure 7. gene-eRNA co-expression network

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A) Graphical presentation of the 153 modules identified by MEGENA. Each node represents a module, the shape of the node indicates the enrichment of SCZ GWAS, node size indicates the number of genes and eRNAs, filled color indicates the fraction of genes within a module, and the circular color indicates the enrichment of SCZ DE genes/eRNAs. Prioritized modes are labeled and highlighted with "+".

B) Prioritized modules enrichment. For the node, the size represents enrichment P value, filled color indicates enrichment fold change, colored circle represents significance level. For SCZ dysregulation and SCZ GWAS enrichment, both gene and eRNA are used. For cell type (Lake et al., 2018), biological process (SynGO and MSigDB) enrichment, gene, and eRNA-linked genes are used. For the development stage (dev), evolutionary conservation enrichment, only eRNA is used.

C) TF enrichment in prioritized modules. The bar plot shows the TF motif overrepresentation at the module eRNA (filled color) and the Spearman correlation with the module eigengene (length).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Supplementary figures

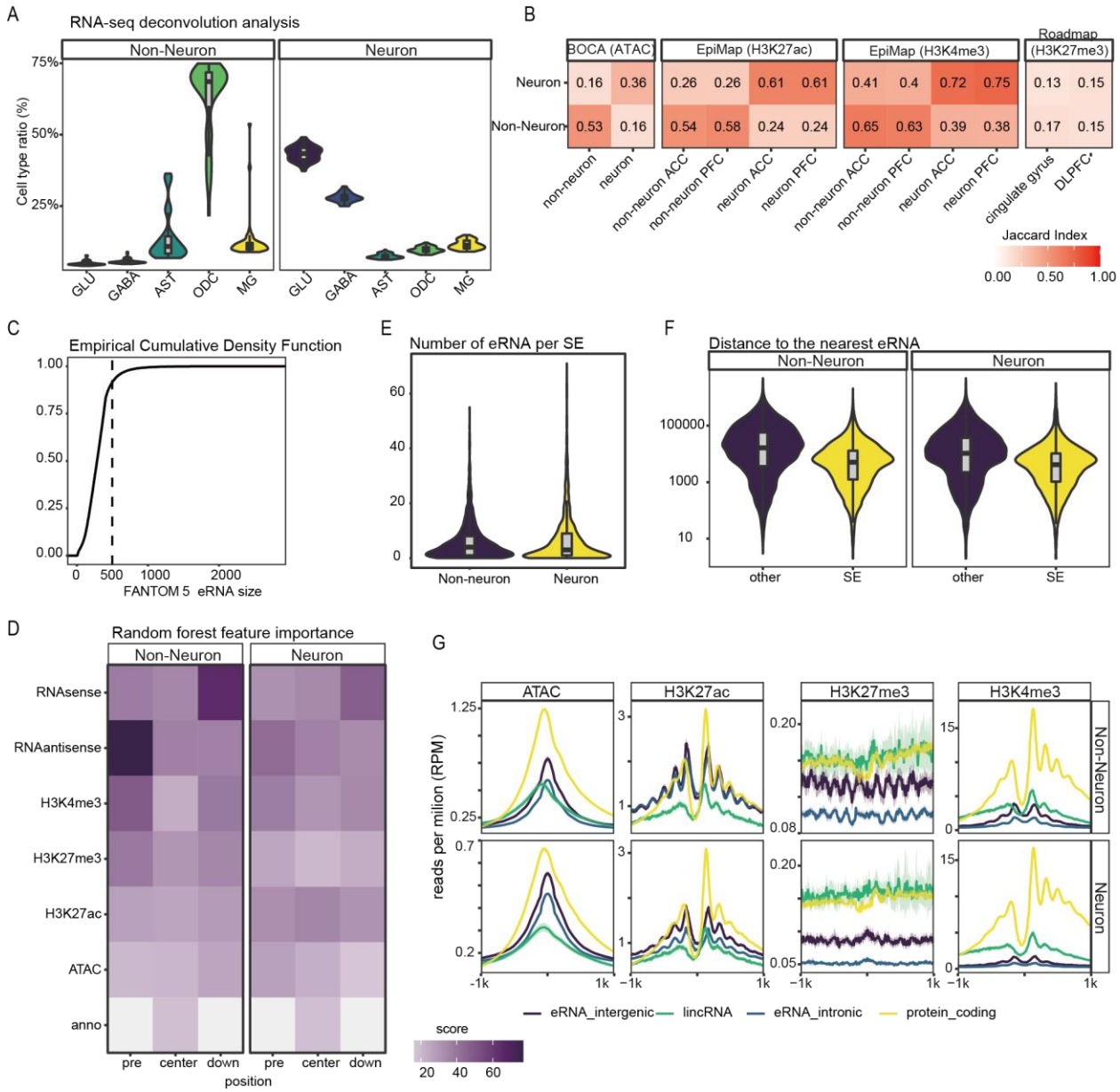


Figure S1 eRNA identification

A) The distribution of deconvolved RNA-seq cell type distribution for each sample.

B) Jaccard index between the Multi-omics peaks and previous reports. cell-type-specific ATAC-seq/H3K4me3/H3K27ac peaks were compared to our previous reports of corresponding assays (Fullard et al., 2018; Girdhar et al., 2018). H3K27me3 was compared to Roadmap H3K27me3 peaks (Roadmap Epigenomics Consortium et al., 2015). PFC indicates the prefrontal cortex.

C) Empirical density distribution of eRNA size from FANTOM5; based on the curve, we chose 500bp as the eRNA size.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

D) Feature importance heatmap from random forest models.
E) Distribution of eRNA numbers per super-enhancer in the two cell types.
F) Distance to the nearest eRNAs for every eRNA within or outside of super-enhancer regions ($p<10^{-16}$ for both cell types, Wilcoxon test).
G) Epigenomic profiles of expressed eRNAs, and promoters of protein-coding genes and lincRNAs. Shadow indicates a 95% confidence interval.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

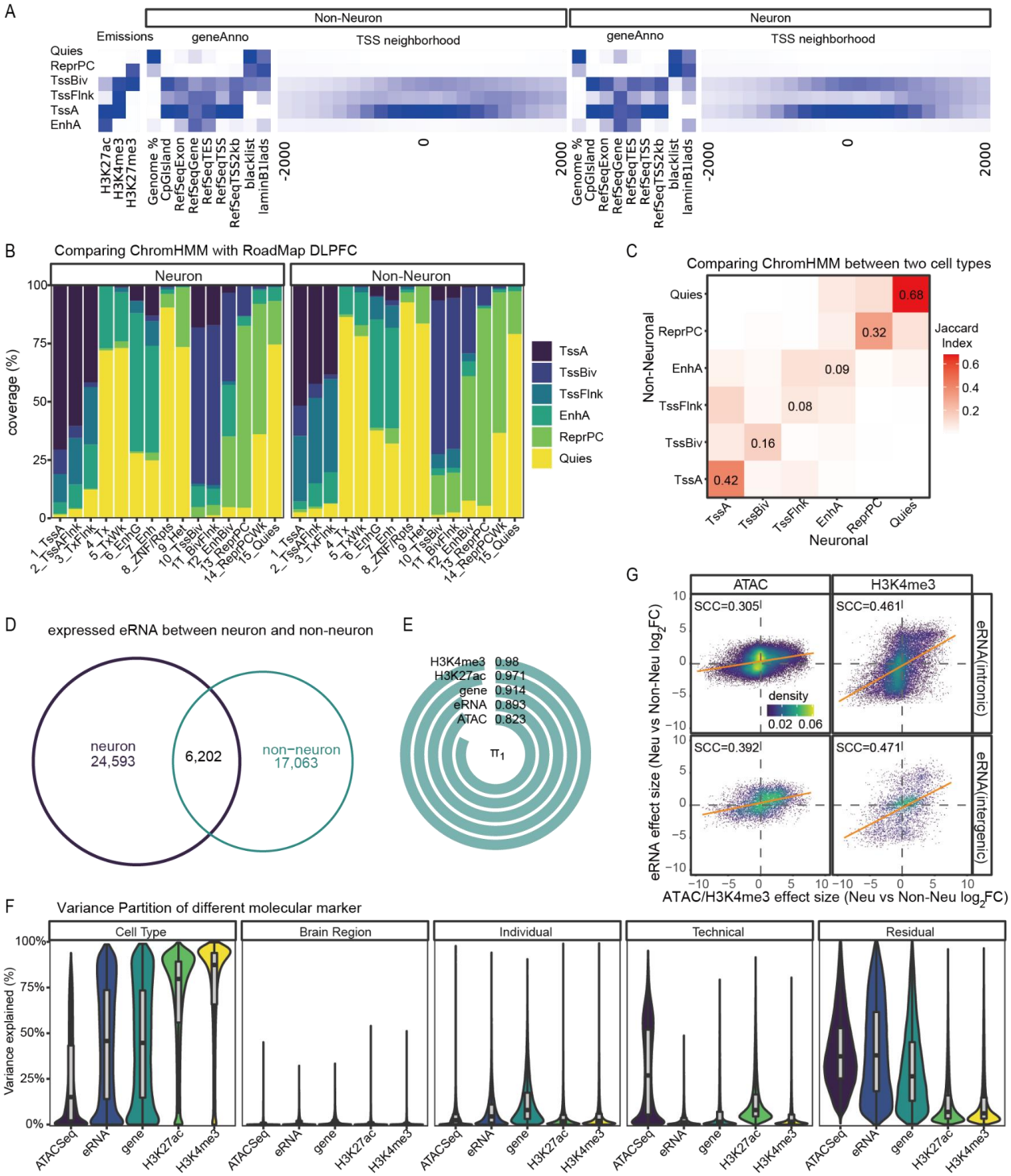


Figure S2 Differential expression/activity between neuronal and non-neuronal cells

A) ChromHMM identified 6 chromatin states including EnhA (active enhancer), TssA (active promoters), TssBiv (bivalent promoters), TssFlnk (promoter flanking region), ReprPC (polycomb repression region), and Quies (other regions).

B) Comparing the cell-type-specific Chromatin states with the roadmap DLPFC result.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C) Jaccard index between neuronal and non-neuronal chromatin states. Compared to active promoters (TssA), and polycomb repressed regions (ReprPC), active enhancers (EnhA) are remarkably more different between cell types.
D) Overlap of expressed eRNAs between neuronal and non-neuronal cells.
E) π_1 statistics of differential activity between the two cell types of different molecular markers.
F) Violin plot shows the variance explained by different factors for the five markers.
G) cell-type-specific effect size ($\log_2$ fold change) between eRNAs and physically overlapping ATAC-seq/H3K4me3 peaks are highly consistent for both intergenic and intronic eRNAs. SCC represents the Spearman correlation coefficient (all $p<10^{-16}$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

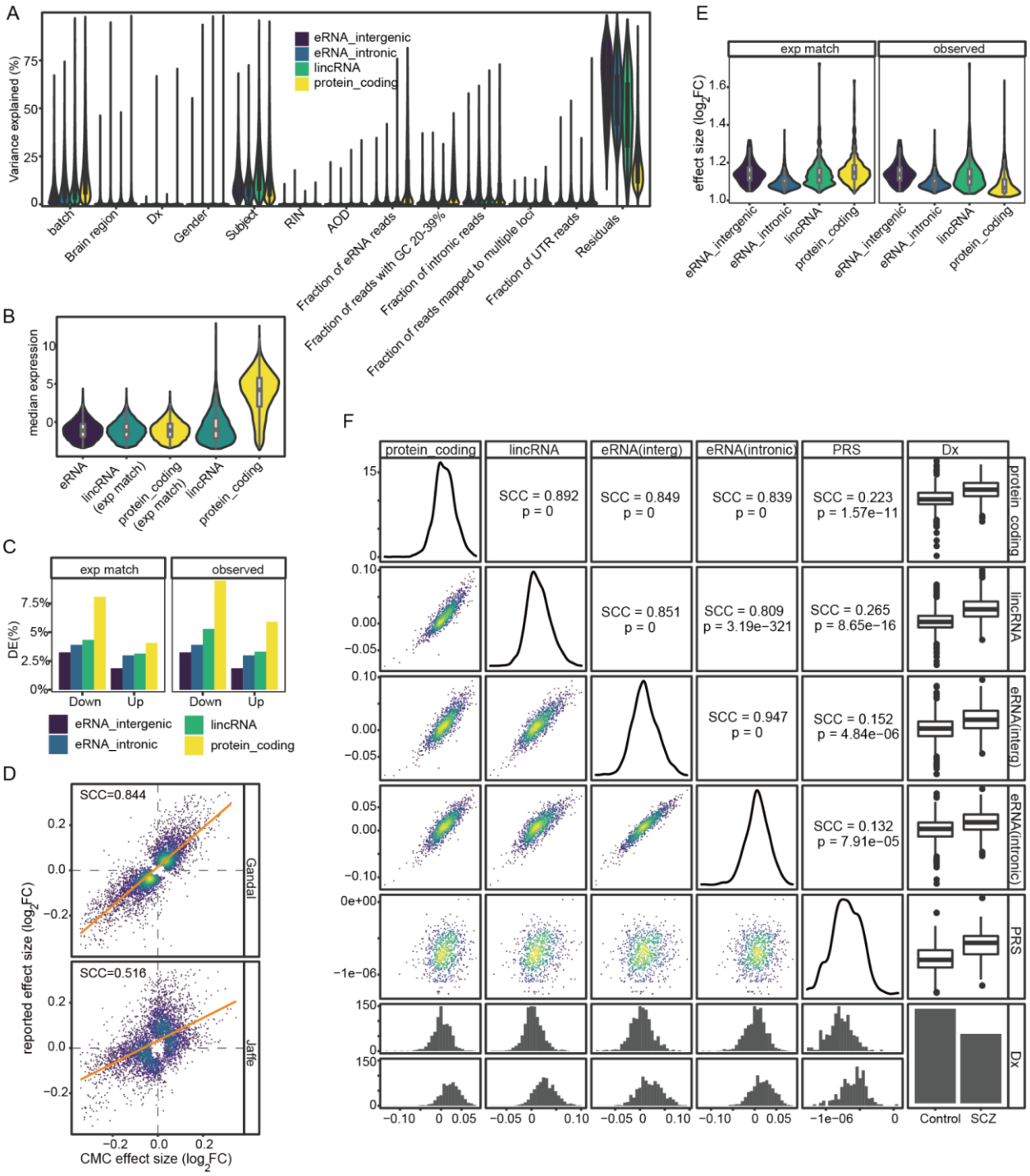


Figure S3 SCZ case-control differential analysis (related Figure 3)

A) Violin plot shows the variance explained by different factors for eRNAs and genes.
B) subsampled the protein-coding gene and lincRNAs to have the same gene expression as that of eRNAs (matched expression).
C) Fraction of differentially expressed transcripts in the observed and expression matched group.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

D) Effect size (log2 fold change) between nominal significant transcripts (p<0.01 in at least one set) are highly consistent ($p<10^{-16}$ for both) between our(CMC) and published results (Gandal et al., 2018; Jaffe et al., 2018). SCC represents the Spearman correlation coefficient.

E) Gene dysregulation fold change between different types of transcripts

F) Correlation between transcriptional perturbation score (summarized by the first principal component of the dysregulated transcriptome) of protein-coding genes, lincRNAs, intergenic eRNA, intronic eRNAs, and PRS and diagnosis(Dx).

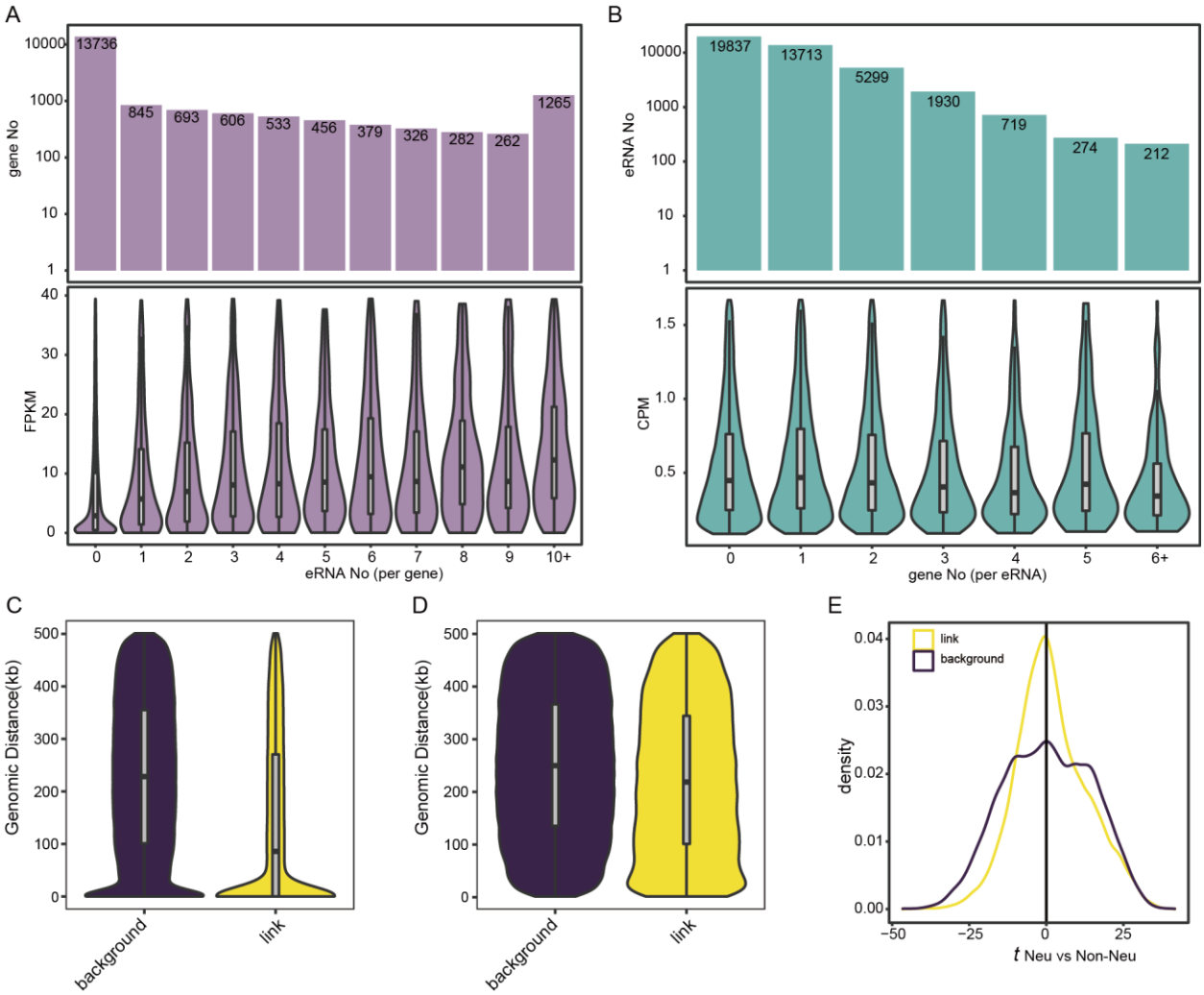


Figure S4 eRNA regulation of target genes (related to Figure4)

A) Distribution of gene FPKM and counts for linked eRNAs. Linked genes have a significantly higher gene expression (FPKM, Wilcoxon test, $p<10^{-16}$).

B) Distribution of eRNA CPM and for counts linked genes. Linked eRNAs do not have a higher expression (CPM, Wilcoxon test, p=0.96).

C) Distance between eRNAs to genes of linked group and background.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

D) Distance between eRNAs to genes of linked group and background, physically-overlapped gene-eRNA pairs were excluded.

E) Neuronal vs non-neuronal *t* statistics for linked and not linked eRNAs. Linked eRNA have a significantly higher value (KS test, $p<10^{-16}$)

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

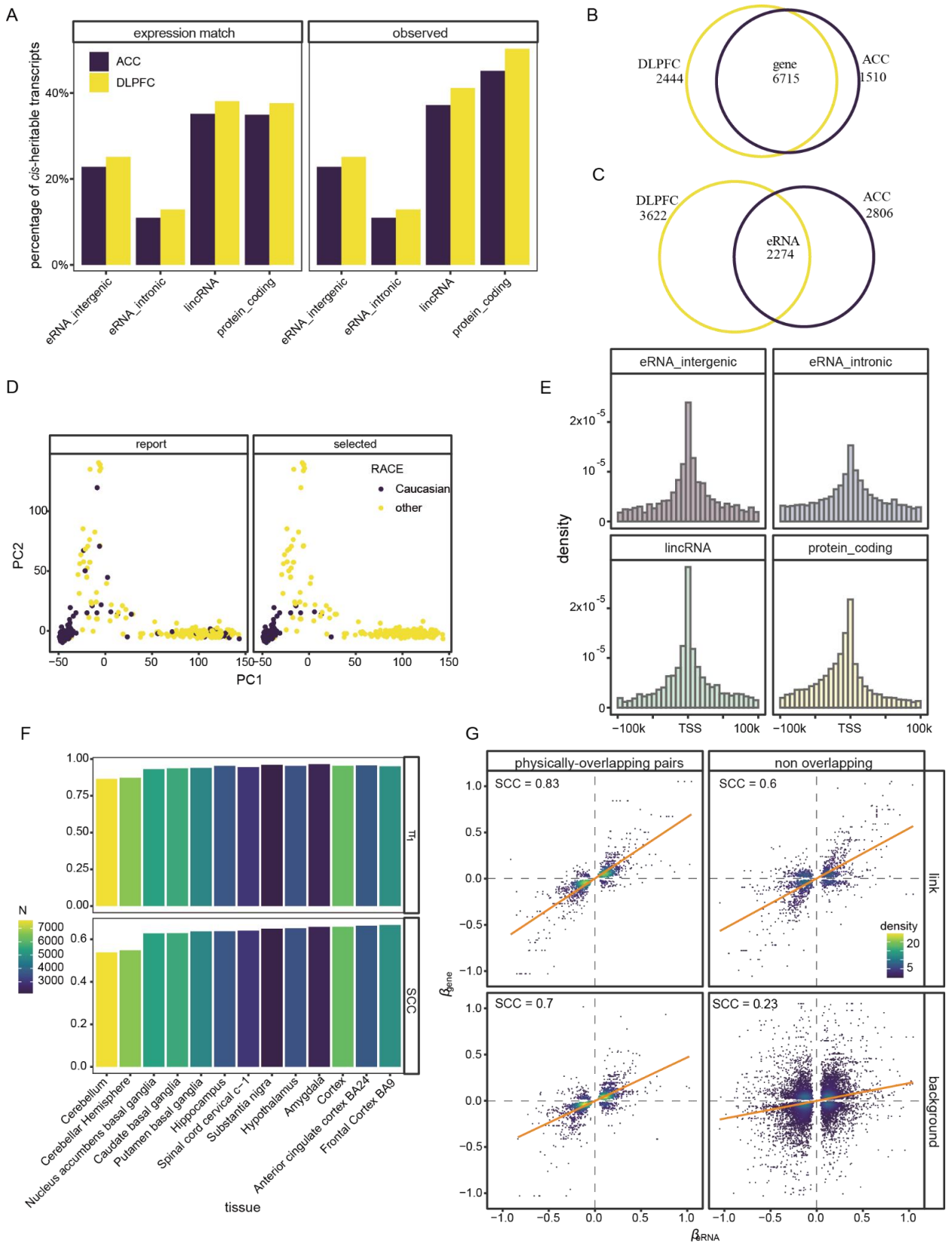


Figure S5 gene and eRNA eQTL

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A) Percentage of expression matched transcripts and observed transcripts that are *cis*-heritable (p<0.05 and *cis*-heritability>0)
B) Overlap of *cis*-heritable genes between ACC and DLPFC
C) Overlap of *cis*-heritable eRNAs between ACC and DLPFC
D) Dot plots illustrate the first two genetic ancestry principal components (PCs) for individuals reported (report) to have European ancestry and the individuals selected (select) to have European ancestry based on sd to the center.
E) The distribution of genomic distances from eSNPs to the TSSs for different classes of transcripts
F) The replication of reported eQTL in our analysis. π_1 values (proportion of true positive p values) for reported significant GTEx eQTL in our gene eQTLs. SCC values are the Spearman correlation coefficients of significant eQTL effect sizes between GTEx eQTL and corresponding pairs in our data. The bar color represents the number of unique genes used.
G) The allelic genetic effect between gene and target eRNAs considering the gene-eRNA physically-overlapping effect.

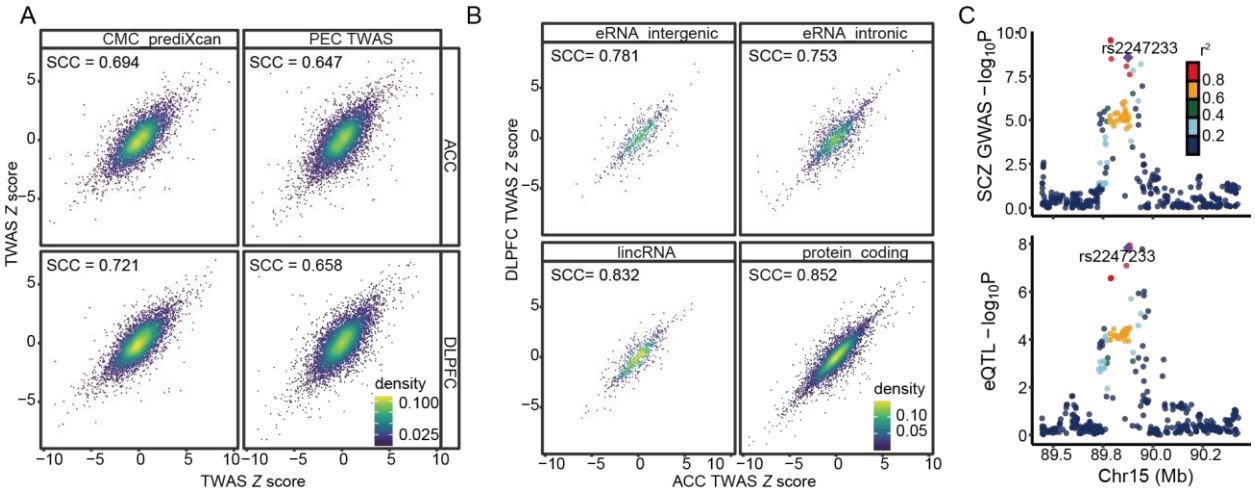


Figure S6 SCZ TWAS

A) Gene TWAS *Z* scores compared to published reports (Gandal et al., 2018; Zhang et al., 2019). SCC represents the Spearman correlation coefficient (ρ)(all $p<10^{-16}$).
B) For different types of transcripts, the TWAS *Z* scores between the two brain regions. SCC represents the Spearman correlation coefficient (ρ)(all $p<10^{-16}$).
C) Aligned Manhattan plots of SCZ GWAS and EeQTLs at the eRNA41216 locus generated by LocusCompare. SNPs are colored by LD (r^2) with the lead EeQTL (rs3247233).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

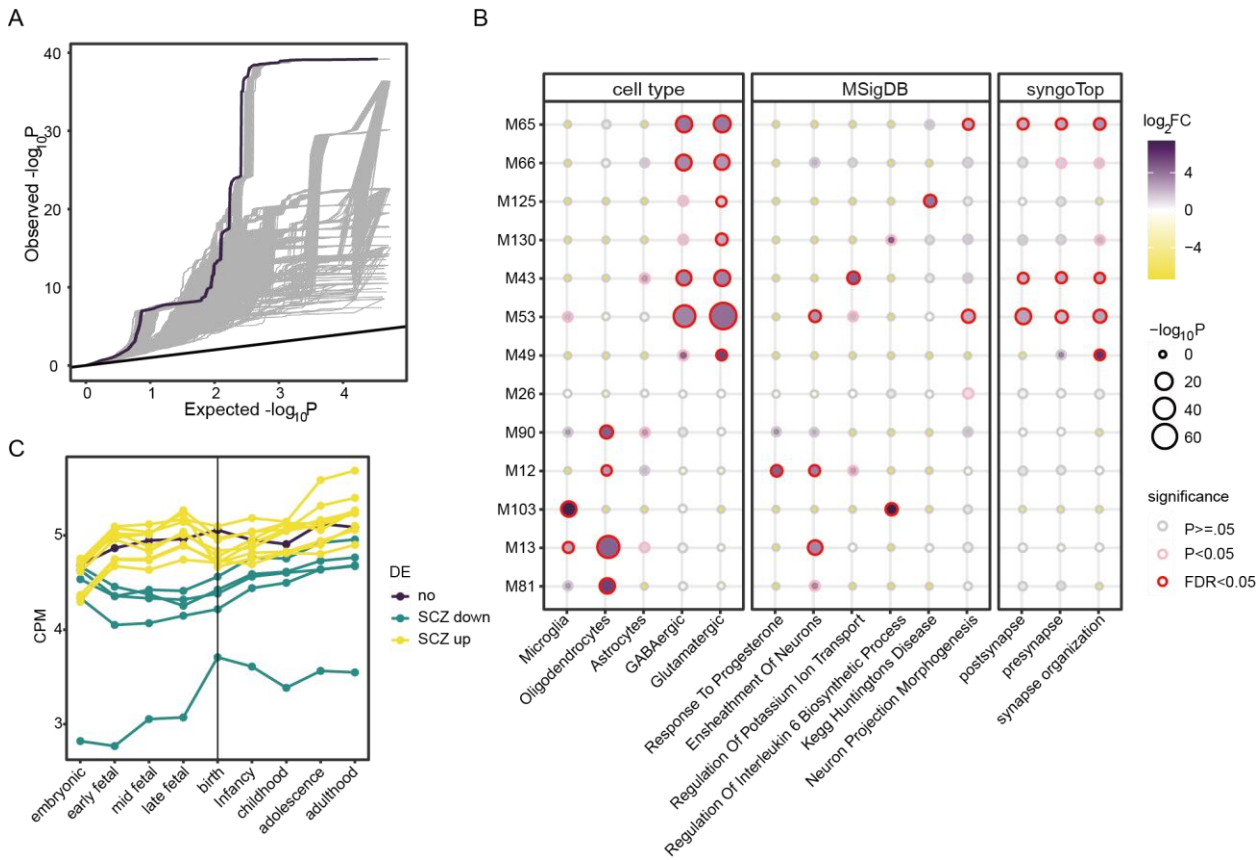


Figure S7 gene-eRNA co-expression network

A) Quantile–quantile plot of SCZ GWAS P value (purple corresponds to M26 eQTL region, and grey represents permutations).
B) Cell types, synaptic GO terms, and biological processes of genes (no eRNA-linked genes) for every module.
C) Median expression of gene and eRNA-linked gene expression in different development stages for every module (Li et al., 2018). Modules are colored by the enriched DE transcripts in SCZ.

Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Panos Roussos (panagiotis.roussos@mssm.edu).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Data and code availability

The clinical information and raw data of Multi-omics cohort including ATAC-Seq, RNA-Seq, H3K4me3/H3K27me3/H3K27ac ChIP-Seq and Hi-C are available via the AD Knowledge Portal (https://adknowledgeportal.org). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.org/DataAccess/Instructions). For access to content described in this manuscript see: http://doi.org/10.7303/syn25672193.

The processed data sets including ATAC-Seq/ChIP-Seq peaks, super-enhancer annotation, chromatin states, eRNA annotation, differential analysis summary statistics, eRNA-gene coordination, eQTL summary statistics, TWAS weights, and module membership, and eRNA identification pipeline are available on the Synapse platform at http://doi.org/10.7303/syn25716684.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Multi-omics human postmortem brain samples

Brain specimens from 10 neurotypical individuals were obtained from the Mount Sinai/JJ Peters VA Medical Center Brain Bank (MSBB–Mount Sinai NIH Neurobiobank), as part of the Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) project (Wang et al., 2018b). All neuropsychological, diagnostic, and autopsy protocols were approved by the Mount Sinai and JJ Peters VA Medical Center Institutional Review Boards. Details about the subject information, including sex, age, postmortem interval (PMI) could be accessed on Synapse via the AD Knowledge Portal (https://adknowledgeportal.org).

CMC human postmortem brain samples

The CMC postmortem brain transcriptomic data consists of two non-overlapping cohorts. The CMC MSSM-Penn-Pitt Cohort (CMC MPP, used as discovery set), consists of samples of Anterior Cingulate Cortex (ACC) and Dorsolateral Prefrontal Cortex (DLPFC) from 545 postmortem brains accessed through the Mount Sinai NIH Brain Bank and Tissue Repository (MSSM), the University of Pennsylvania Brain Bank of Psychiatric illnesses and Alzheimer's Disease Core Center (Penn), and the University of Pittsburgh NIH NeuroBioBank Brain and Tissue Repository (Pitt). The CMC HBCC Cohort (CMC_HBCC, used as a replicate set), consists of samples (ACC and DLPFC) from 229 postmortem brains accessed through the NIMH Human Brain Collection Core (HBCC). For both cohorts, dissections were performed at the corresponding brain banks and sent to Icahn School of Medicine at Mount Sinai for sample processing and library preparation. Postmortem tissue from schizophrenia (SCZ) cases was included if they met the diagnostic criteria in DSM-IV

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

for SCZ or schizoaffective disorder after review of medical records, direct clinical assessments, and interviews with care providers. Individuals that had been diagnosed with, or had a history of other neuropsychiatric disorders (including bipolar disorder, and/or Parkinson's disease), or acute neurological insults (anoxia, strokes, and/or traumatic brain injury), or Alzheimer's disease were excluded from this analysis. Additionally, we excluded samples with age at death <17 and/or, PMI>48. All neuropsychological, diagnostic, and autopsy protocols were approved by the Mount Sinai and JJ Peters VA Medical Center Institutional Review Boards. The detailed subject information can be accessed through CMC Knowledge Portal (http://CommonMind.org).

CRISPR subject information

Two control individuals NSB553-S1-1 (male), and NSB2607-1-4 (male) were selected from a previously reported case-control hiPSC childhood-onset SCZ cohort for CRISPR interference experimental validation (Hoffman et al., 2017; Schrode et al., 2019). This work was reviewed by the Internal Review Board of the Icahn School of Medicine at Mount Sinai, and the Embryonic Stem Cell Research Oversight Committee at the Icahn School of Medicine at Mount Sinai.

METHOD DETAILS

Molecular Data Generation (Multi-omic Cohort)

FANS sorting of neuronal and non-neuronal nuclei

50 mg of frozen brain tissue was homogenized in chilled lysis buffer (0.32 M Sucrose, 5 mM $CaCl_2$, 3 mM Magnesium acetate, 0.1 mM, EDTA, 10 mM Tris-HCl, pH 8, 1 mM DTT, 0.1% Triton X-100) and filtered through a 40 µm cell strainer. Filtered lysate was underlaid with sucrose solution (1.8 M Sucrose, 3 mM $Mg(CH_3COO)_2$, 1 mM DTT, 10 mM Tris-HCl, pH 8) and subjected to ultracentrifugation at 107,000 xg for 1 hour at 4°C. Pellets were resuspended in 500 µl DPBS containing 0.1% BSA. anti-NeuN antibody (1:1000, Alexa488 conjugated, Millipore, Cat# MAB377X) was added and samples incubated, in the dark, for 1 hr at 4°C. Prior to FANS sorting, DAPI (Thermoscientific) was added to a final concentration of 1 µg/ml. DAPI positive neuronal (NeuN+) and non-neuronal (NeuN-) nuclei were isolated using a FACSAria flow cytometer with FACSDiva Version 8.0.1 software (BD Biosciences).

Generation of RNA-seq libraries and sequencing

For RNA-seq, nuclei were sorted into 1.5 ml low-binding Eppendorf tubes containing Extraction buffer, a component of the PicoPure RNA Extraction kit (Arcturus, cat.# KIT0204). RNA was isolated following the manufacturer's instructions. This included an RNase-free DNase treatment step (Qiagen, cat.# 79254). Samples were eluted in RNase-free water and stored at -80°C until preparation of RNA-sequencing libraries using the SMARTer Stranded Total RNA-seq Pico Kit v1 (Takara Clontech Laboratories, cat.# 635005) according to the manufacturer's instructions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Following the construction of the RNA-seq libraries, libraries were quantified by quantitative PCR (KAPA Biosystems, cat.# KK4873) and library fragment sizes estimated using High Sensitivity Tapestation D1000 ScreenTapes (Agilent, cat.# 5067-5584). RNA-seq libraries were subsequently sequenced on Hi-Seq2500 (Illumina) machines yielding 150 bp paired-end reads.

Generation of ATAC-seq libraries and sequencing

ATAC-seq libraries were generated using an established protocol (Buenrostro et al., 2015). Briefly, 55,000 to 75,000 sorted nuclei were pelleted at 500 xg for 10 min at 4°C. Pellets were resuspended in transposase reaction mix (22.5 µL Nuclease Free H_2O, 25 µL 2x TD Buffer; Illumina, cat.# FC-121-1030) and 2.5 µL Tn5 Transposase (Illumina, cat.# FC-121-1030) on ice and the reactions incubated at 37 °C for 30 min. Following incubation, samples were purified using the MinElute Reaction Cleanup kit (Qiagen cat.# 28204), and libraries generated using the Nextera index kit (Illumina cat. FC-121-1011). Following amplification, libraries were resolved on 2% agarose gels and fragments ranging in size from 100-1000 bp were excised and purified (Qiagen Minelute Gel Extraction Kit, Qiagen, cat.# 28604). Next, libraries were quantified by quantitative PCR (KAPA Biosystems, cat.# KK4873) and library fragment sizes estimated using Tapestation D5000 ScreenTapes (Agilent technologie, cat. 5067-5588). ATAC-seq libraries were subsequently sequenced on Hi-Seq2500 (Illumina) machines yielding 50 bp paired-end reads.

Generation of ChIP-seq libraries and sequencing

Ultra-low input Native Chromatin Immunoprecipitation sequencing (ULI-NChIP-seq) libraries were prepared as follows. After isolating neuronal (NeuN+) and non-neuronal nuclei (NeuN- nuclei) by FANS, we performed ULI-NChIP assays, adapted from Brind'Amour, et al., 2015, which specifically does not require chromatin crosslinking, thereby increasing library complexity and reducing PCR artifacts. Briefly, nuclei were centrifuged at 500xg for 10 minutes at 4°C and re-suspended by gentle pipetting in residual smaller volumes of PBS/Sheath buffer. After DAPI counting of nuclei (Countess II, Life technologies), 100K-300K was distributed into Eppendorf tubes and 0.1% Triton-X-100/0.1% Na-Deoxycholate added. The chromatin was re-suspended and placed at room temperature for 5 minutes, followed by fragmentation with micrococcal nuclease (MNase, NEB, cat.# M0247S) for 5 min at 37°C on a ThermoShaker at 800 rpm to digest the chromatin to, predominantly, mononucleosomes. The MNase reaction was stopped by addition of 10% of the reaction volume of 100 mM EDTA (pipetted ~20x), followed by the addition of 1% Triton / 1% deoxycholate, (pipetted 5x) and placed on ice for at least 15 minutes. Samples were vortexed (medium setting) for ~ 30 seconds and complete NChIP buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 15 mM NaCl, 0.1% Triton X-100, 1 EDTA-free protease inhibitor cocktail and 1 mM phenylmethanesulfonyl fluoride) was added to dilute the chromatin to <25% of the immunoprecipitation reaction volume, and rotated at 4°C for 1 hour. After the incubation, chromatin was vortexed (medium setting) for ~30 seconds and 5% input controls removed for DNA extraction.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To avoid non-specific binding, chromatin was next precleared by adding 10 µl/ reaction of a pre-washed 1:1 ratio of protein A to protein G Dynabeads (Life Technologies), rotating the chromatin-protein A/ protein G magnetic bead mixture for 3 hours at 4°C. Antibody-bead complexes were prepared as follows: ChIP-grade histone H3K27ac antibodies (Active Motif, cat.# 39133, pAB), H3K4me3 (Cell Signaling, cat.# 9751), or H3K27me3 (Cell Signaling, cat.# 9733) were added to prewashed protein A/protein G Dynabeads resuspended in NChIP buffer, and the antibody-bead complexes formed by rotating the antibody and beads for 2 hours at 4°C. After the incubations, precleared chromatin and antibody-bead complexes were placed on a magnetic rack and the precleared chromatin transferred to new Eppendorf tubes, while the antibody-bead complexes were resuspended in sufficient volume of NChIP buffer to add 10 µl per MNase reaction. The chromatin was immunoprecipitated with the antibody-bead complexes at 4°C overnight while rotating.

After the overnight incubation, the immunoprecipitation reactions were placed on a magnetic rack to remove unbound chromatin, washed twice with 200 µl of ChIP low salt wash buffer (20 mM Tris-HCl, pH 8.0, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 2 mM EDTA and 150 mM NaCl), twice with 200 µl ChIP high salt wash buffer (20 mM Tris-HCl (pH 8.0), 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 2 mM EDTA and 500 mM NaCl), followed by elution in freshly prepared 30 µl of ChIP elution (100 mM sodium bicarbonate and 1% SDS) for 1.5 hours at 68°C on a ThermoShaker at 1000 rpm, RNAse A digestion for 15 minutes at 37˚C at 800 rpm. The immunoprecipitated DNA, along with the input controls, was purified using Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) (ThermoFisher Scientific, cat.# 15593-031), transferred to pre-spun phase lock tubes (Qiagen Maxtract, cat.# 129046) to obtain the aqueous layer. An overnight ethanol precipitation was performed by adding 10 µl of 3M sodium acetate/100 µl aqueous layer of ChIP DNA, 1 µl of LPA (linear polyacrylamide, Sigma #56575) and 1 µl Glycoblue (Invitrogen, cat.# AM9515) in 275 µl 100% ethanol.

NEBNext Ultra DNA Library Prep Kit (New England Biolabs, E7370) was used to construct NChIP libraries according to the manufacturer's directions, followed by Pippin Size selection using 2% Agarose Gel cassettes (SAGE Science, cat.# HTC2010) and clean up with 1.8 volumes of SPRIselect beads (Beckman Coulter, cat.# B23317). All libraries were analyzed on an Agilent High Sensitivity D1000 TapeStation, and quantified using the KAPA Library Quantification Kit prior to sequencing. The constructed libraries, each containing a unique index, were pooled and sequenced using the Nova-seq platform (Illumina).

Generation of Hi-C libraries and sequencing

Hi-C data was generated from frozen postmortem human brain tissue using the in situ Hi-C protocol (Rao et al., 2014) with the following modifications. Frozen brain tissue was thawed at room temperature (RT) and dounce homogenized in HBSS (Hank's balanced salt solution). Homogenized tissue was fixed with 0.5% formaldehyde for 10 min and then quenched with 0.125 M glycine for 5 min at RT. Cross-linked tissue was then placed on ice for a further 15 min to quench crosslinking completely. Samples were centrifuged at 800 xg for 10 min at 4°C and pellets resuspended in lysis buffer (0.32 M Sucrose, 5 mM $CaCl_2$, 3 mM $Mg(CH_3COO)_2$, 0.1 mM EDTA, 10 mM Tris-HCl, pH 8, 1 mM DTT, 0.1 % Triton X-100, 1x Roche cOmplete mini EDTA-free

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

protease inhibitor tablet (Roche cat.# 4693159001) to isolate cross-linked nuclei. Neuronal nuclei were FANS sorted using the Alexa488 conjugated anti-NeuN antibody (1:1000) (Millipore, cat.# MAB377X) using a BD FACS Aria II sorter. FANS sorted neuronal (NeuN+) and non-neuronal (NeuN-) nuclei were pelleted at 2500 xg for 5 min at 4°C, frozen on dry ice for 20 min, and then stored at -80°C.

Approximately 1 M crosslinked neuronal and non-neuronal nuclei were thawed on ice, washed with ice-cold 1x CutSmart buffer (New England Biolabs (NEB), cat. B7204S) and split into 4 aliquots to generate technical replicate libraries per sample, with 250k nuclei per library. Nuclei were pelleted at 2500 xg for 5 min at 4°C, and resuspended in 342 µl 1x CutSmart buffer and then conditioned with 0.1% SDS at 65°C for 10 min. Nuclei were immediately placed on ice and the SDS quenched with 1% Triton X-100. Chromatin was digested with 100 U of the 4 base pair cutter MboI (NEB, cat.# R0147L) overnight at 37°C with shaking at 400 rpm. MboI was heat-inactivated at 65°C for 20 min, and then nuclei were cooled down on ice. MboI cut sites were end-labeled with biotin by adding 52 µl of biotin fill-in reaction mix (15 µl of 1 mM biotin-14-dATP (Jena Bioscience, cat.# NU-835-BIO14-L), 1.5 µl each of 10 mM dCTP, dGTP, dTTP (Sigma-Aldrich, cat.# DNTP10-1KT), 10 µl of 5 U/µl Klenow DNA Pol I (NEB, cat.# M0210L), 22.5 µl of 1x CutSmart buffer) and shaking at 37°C for 1.5 h at 400 rpm. Blunt ended sites were proximity ligated by adding 948 µl ligation reaction mix (150 µl of 10x T4 DNA ligase buffer (NEB, cat.# B0202S), 125 µl of 10% Triton X-100, 15 µl of 10 mg/ml BSA, 10 µl of 400 U/µl T4 DNA ligase (NEB, cat.# M0202L), 648 µl ddH_20) and rotating tubes, end-over-end, at RT for 4 h. Nuclei were reverse crosslinked with 100 µl proteinase K (10 mg/ml) overnight at 65°C.

Proximity ligated DNA was purified through phenol:chloroform extraction and sodium acetate/ethanol precipitation. Purified DNA was sheared using a Covaris S220 sonicator to generate a peak size of 400 bp with the following settings (peak incident power: 140 W, duty cycle: 10%, cycles per burst: 200, time: 55 sec). Biotin-labeled ligation junctions were purified with Dynabeads MyOne Streptavidin C1 beads (ThermoFisher, cat.# 65001) by incubating for 1 hr at RT. Illumina compatible libraries were prepared from the sonicated and streptavidin bead immobilized DNA using the NEBNext Ultra II Library prep kit (NEB, cat.# E7645L), following manufacturer's instructions, by amplifying libraries for 6-10 PCR cycles. Libraries were purified by 2-sided size selection (300-800 bp) using Ampure XP beads (Beckman Coulter, cat.# A63881). All libraries were analyzed on a TapeStation using Agilent D5000 ScreenTapes (Agilent technologies, cat.# 5067-5588), and quantified using the KAPA Library Quantification Kit (KAPA Biosystems, cat.# KK4873) prior to sequencing. Uniquely barcoded Hi-C libraries were pooled and deep sequenced on the Illumina NovaSeq S4 platform (Illumina) obtaining 100 bp paired-end reads.

Molecular Data Generation (CMC Cohort)

Generation of RNA-seq libraries

CMC MPP cohorts. Briefly, rRNA was depleted from about 1 µg of total RNA using Ribozero Magnetic Gold kit (Illumina/Epicenter, cat.# MRZG12324) to enrich polyadenylated coding RNA

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and non-coding RNA. The sequencing library was prepared using the TruSeq RNA Sample Preparation Kit v2 (Illumina, cat.# RS-122-2001). The insert size and DNA concentration of the sequencing library were determined on Agilent Bioanalyzer and Qubit, respectively. A pool of 10 barcoded libraries was layered on a random selection of two of the eight lanes of the Illumina flow cell at appropriate concentration and bridge amplified to ~250 million raw clusters. 100bp paired-end reads were obtained on a HiSeq 2500 (Illumina).

HBCC cohort. Briefly, rRNA was depleted from 1ug of RNA using the KAPA RiboErase protocol that is integrated into the KAPA Stranded RNA-seq Kit (KAPA Biosystems, cat.# KK8483). The insert size and DNA concentration of the sequencing library were determined on Fragment Analyzer Automated CE System (Advanced Analytical) and Quant-iT PicoGreen (ThermoFisher, cat.# P7589), respectively. A pool of 10 barcoded libraries was layered on a random selection of two of the eight lanes of the Illumina flow cell at appropriate concentration and bridge amplified to ~250 million raw clusters. 100bp paired-end reads were obtained by HiSeq 2500 (Illumina).

CRISPR inference validation

Cell culture and transduction of dCas9-effector hiPSC-NPCs with gRNA lentivirus. The iPSC lines used in this study (NSB2607-1-4 and NSB553-S1-1) (Schrode et al., 2019) were maintained in Matrigel (Corning, cat.# 354230) coated 6-well plates under NPC medium (DMEM/F12 (Life Technologies, cat.# 10565), 1x N2 (Life Technologies, cat.# 17502-048), 1x B27-RA (Life Technologies, cat.# 12587-010), 20 ng/ml FGF2 (R&D Systems, cat.# 233-FB-01M). Upon confluency, cells were dissociated with Accutase (Innovative Cell Technologies, cat.# AT104) for 5 minutes at 37°C, quenched with DMEM/F12, pelleted, and resuspended in NPC medium containing 10µM/ml Thiazovivin (THX) (Sigma/Millipore, cat.# 420220). 3.5x10^5 NPCs per well were seeded onto Matrigel-coated 24-well plates in NPC media. The following day, gRNA lentiviruses such as LentiGuide-Hygro-mTagBFP2 (Addgene, cat.# 99374) were added to cultures, followed by spinfection (1 hour, 1000xg, 25°C). After spinfection, the cultures were incubated overnight, and medium was then replaced the following day. The cells were selected with 0.3 µg/ml puromycin for dCas9-KRAB (Sigma, cat.# P7255) for two days. After puromycin selection, cells were fed with fresh NPC medium and then harvested two days later. gRNA expression was confirmed via BFP fluorescence, prior to harvest.

The Benchling CRISPR gRNA design tool was used to design guide sequences. Two guide RNAs for eRNA41216 with the highest specificity scores were selected:

Guide RNA1 (chr15: 89400303): 5'-GAGGCGCGATACGAACCCGT-3'

Guide RNA2 (chr15: 89400387): 5'-GATACGGGCGAATCCCGCAA-3'

The guide oligos were then synthesized by IDT (Integrated DNA technologies) and phospho annealed via T4 PNK (37°C for 30min, 95C for 5min and ramped down to 25°C at 5°C/min). The annealed oligos were then cloned downstream of a constitutively expressed U6 promoter in a lentiviral vector (lentiGuide-Hygro-mTagBFP2, Addgene, cat.# 99374) using the golden gate cloning method (Bsmb1; 30 cycles of 37°C for 5min and 20°C for 5min). Vectors were packaged into 3rd generation lenti-viruses by VectorBuilder and transduced into dCas9-KRAB expressing neural progenitor cells (NPCs). RNA was extracted from transduced NPCs and target gene expression knockdown was validated through TaqMan RT-qPCR relative to cells that were

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

transduced with a negative control scrambled guide. The 2–ΔΔCt method was used to determine fold change in expression relative to the housekeeping genes GAPDH and ACTB.

Molecular Data Processing (Multi-omic Cohort)

We checked the matched fastq files by Pool ID and barcodes matching and checked for MD5 file integrity checksums after receiving sequencing data. The checked fastq files were trimmed with Trimmomatic (v0.36) (Bolger et al., 2014) for further processing.

Whole-genome Sequencing

Whole-genome sequencing (WGS) data were obtained from AMP-AD MSBB cohort (Wang et al., 2018b). The paired-end 150 bp reads were aligned to the human hg38 reference genome with Burrows–Wheeler Aligner (bwa) using the bwa-mem algorithm (Li, 2013). Duplicates were marked and discarded by Picard. Then the bam files were realigned around indels, and base recalibrated with GATK. For each individual, variants were called using GATK HaplotypeCaller. The individual files were then jointly genotyped using GATK GenotypeGVCFs and generated a multi-sample VCF. Finally, quality metrics for each variant were determined through Variant Quality Score Recalibration (VQSR).

RNA-seq data processing

The trimmed reads were aligned to human genome hg38 (GRCh38) using STAR (2.7.2a) aligner (Dobin et al., 2013), where the allelic alignment bias was corrected by WASP (van de Geijn et al., 2015). Gene expression was quantified using RSEM (v1.3.1) tools (Li and Dewey, 2011) with GENCODE V30 as a reference and summarized at the gene level. Duplication and GC content levels were estimated by Picard tools (v2.2.4). Quality control (QC) metrics were collected by RNA-seq QC (v1.1.8) (DeLuca et al., 2012). The RNA-seq libraries were deeply sequenced and have acceptable values for total reads (mean 1.82×10^8, sd $\pm 9.5 \times 10^7$), final read pairs (mean 1.48×10^8 sd $\pm 7.44 \times 10^7$), intergenic rate (mean 11.7%, sd ±1.86%), intronic rate (mean 59.3%, sd ±3.38%), and ribosomal RNA rate (mean 0.0906%, sd ±0.0631%). We checked the sex compatibility using the expression of two sex-specific genes XIST and RPS4Y1.

RNA-Seq cell type deconvolution analysis. We used dTangle (v2.0.9) (Hunt et al., 2019), a method built on the linear mixing model of linear-scale expressions of known marker genes, to estimate the cell type composition of RNA-seq data. Human brain single-cell markers including five brain cell types: glutamatergic neurons, GABAergic neurons, astrocytes, oligodendrocytes, and microglia were used as the reference (Lake et al., 2018). As expected, the neuronal cells are enriched for GABAergic and glutamatergic, while the non-neuronal cells are strongly enriched for oligodendrocytes and followed by astrocytes and microglia (**Figure S1A**).

ATAC-seq data processing

Alignment. The trimmed reads were aligned to human genome hg38 (GRCh38) using STAR (2.7.0e) aligner (Dobin et al., 2013), with customized parameters. We restricted the intron length (--alignIntronMax 1), number of mismatches per pair (--outFilterMismatchNmax 100), alignment score (--outFilterScoreMinOverLread 0.66) and number of matched bases (--outFilterMatchNminOverLread 0.66) in local alignment mode (-alignEndsType Local), and

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

controlled allelic alignment bias with WASP (van de Geijn et al., 2015). Reads mapped to multiple loci detected by samtools (v0.1.19), duplicated reads that marked by Picard (v2.2.4), and mitochondria alignments were filtered.

Peak calling. We called peaks for each cell type independently. For each cell type, bam files from each sample were subsampled to the same sequencing depth and then merged. We used model-based Analysis of ChIP-seq (MACS, v2.1) (Zhang et al., 2008) to call peaks with a smoothing window of 200bps (--shift -100 --extsize 200 --nomodel) and FDR of 0.01 (-q 0.01). Peaks overlapped with ENCODE blacklist regions (Amemiya et al., 2019) were discarded, yielding 222,746 neuronal and 136,171 non-neuronal OCRs. To confirm our results, we determined the Jaccard index between our peaks and previously published results (Fullard et al., 2018), which is the intersection of base pairs divided by the union of base pairs (**Figure S1B**).

Quantification. A consensus peak set of both cell types was generated for quantification and differential analysis. We counted the number of reads that overlapped with the consensus peaks using featureCounts function in RSubread (v 1.6.3) (Liao et al., 2014).

QC. We collected the following information for each sample for QC. 1) total number of initial reads; 2) the number of uniquely mapped reads; 3) the fraction of reads that were uniquely mapped; 4) further mappability-related metrics from the STAR aligner; 5) GC content, 6) insert and duplication metrics from Picard; 6) the rate of reads mapping to the mitochondrial genome; 7) the PCR bottleneck coefficient (PBC), which approximates library complexity as uniquely mapped non-redundant reads, divided by the number uniquely mapped reads; 8) the relative strand cross-correlation coefficient (RSC) and the normalized strand cross-correlation coefficient (NSC), which are metrics that use cross-correlation of stranded read density profiles to evaluate the sample quality independently of peak calling; and, 10) finally, the fraction of reads in peaks (FRiP), which is the fraction of reads that fall in called peaks within backlist regions or not. The ATAC-seq libraries have acceptable values for: total reads (mean 5.61×10^7, sd ± 9.2×10^6), final read pairs (mean 3.63×10^7, sd ± 8.07×10^6), mitochondrial alignment rate (mean 1.59%, sd ± 0.61%), fraction of reads in non-blacklist peaks (mean 17.6%, sd ± 4.12%), and mean GC content (mean 51.4%, sd ± 3.58%).

ChIP-seq data processing

H3K4me3, H3K27ac, H3K27me3 ChIP-seq, and corresponding input files were aligned to human genome hg38, similar to our ATAC-seq pipeline, above. The resulting bam files were subsampled to the same sequencing depth and merged for each cell type. We called peaks using MACS with a smoothing window of 150bps (--shift -75 --extsize 150 --nomodel), FDR of 0.01 (-q 0.01), and input as control. For H3K27me3, we called broad peaks (--broad). Peaks overlapped with ENCODE blacklist regions (Amemiya et al., 2019) were discarded, yielding 76,905 H3K4me3, 312,658 H3K27ac, and 83,865 H3K27me3 neuronal peaks, as well as 96,015 H3K4me3, 227,155 H3K27ac, and 68,457 H3K27me3 non-neuronal peaks. Comparing H3K4me3 and H3K27ac peaks to reported peaks (Girdhar et al., 2018) by determining the Jaccard index (**Figure S1B**), we observed high concordance between cell types. For H3K27me3 peaks, we compared our cell-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

type-specific peaks with brain DLPFC bulk H3K27me3 peaks (Roadmap Epigenomics Consortium et al., 2015). A consensus peak set was generated for quantification and differential analysis. We quantified the number of reads overlapped with the consensus peaks using featureCounts function in RSubread (v 1.6.3) (Liao et al., 2014). The QC metrics were collected as for the ATAC-seq data, above.The H3K4me3 ChIP-seq libraries have acceptable values for: total read (mean $1.18x10^8$, sd ± $3.6x10^7$), final read pairs (mean $5.98x10^7$, sd ± $1.87x10^7$), fraction of reads in non-blacklist peaks (mean 83.1%, sd ± 9.23%), NSC (mean 2.14, sd ± 0.24), and RSC (mean 1.46, sd ± 0.135). The H3K27ac ChIP-seq libraries have acceptable values for: total read (mean $1.04x10^8$, sd ± $2.45x10^7$), final read pairs (mean $7.23x10^7$, sd ± $1.91x10^7$), fraction of reads in non-blacklist peaks (mean 71.5%, sd ± 5.76%), NSC (mean 1.19, sd ± 0.0425), and RSC (mean 3.25, sd ± 0.748). The H3K27me3 ChIP-seq libraries have acceptable values for: total read (mean $1.78x10^8$, sd ± $8.03x10^7$), final read pairs (mean $1.35x10^8$, sd ± $5.04x10^7$), fraction of reads in non-blacklist peaks (mean 44.2%, sd ± 1.07%), NSC (mean 1.09, sd ± 0.0377), and RSC (mean 3.09, sd ± 1.09).

Super enhancer identification

We identified super-enhancers with H3K27ac ChIP-seq peaks and bam files using the ROSE pipeline (Whyte et al., 2013). Briefly, enhancer peaks were stitched together if they are located within 12.5 kb of each other and don't have multiple active promoters in between. The stitched peaks were then ranked according to increasing H3K27ac signal intensity. We identified 2,049 neuronal and 1,946 non-neuronal super-enhancers, respectively.

Chromatin states annotation

We implemented a multivariate Hidden Markov Model model (ChromHMM) (Ernst and Kellis, 2012) to systematically annotate the combinational effect of different histone modifications. The ChromHMM model is trained by virtually concatenating histone marks H3K4me3, H3K27ac, and H3K27me3 in both cell types that merged across all individuals that subsampled to uniform depths. Reads were shifted from 5' to 3' direction by 100 bp for all the samples. Read counts were then computed in 200bp non-overlapping bins across the genome. Each bin was binarized into 1 or 0 by the Poisson model with a p-value threshold of 10^{-4}. We have trained the model with merged data using six states which captured all the key interactions from our data. Lastly, we obtained the chromatin states with the trained model and corresponding binarized files as input (**Figure S2A**).

Our models agreed well with the published brain DLPFC region 11 chromatin states model (Roadmap Epigenomics Consortium et al., 2015) (**Figure S2B**).

Hi-C data analysis

Hi-C data were aligned using the HiC-Pro strategy (Servant et al., 2015). Briefly, paired-end reads were mapped independently to the human genome hg38 using bowtie2 in stringent mode with parameters ("--very-sensitive -L 20 --score-min L,-0.6,-0.2 --end-to-end") (Langmead and Salzberg, 2012). Then, the chimeric reads that failed to align were trimmed after ligation sites (MboI "GATCGATC") and mapped to the genome. All the aligned reads from both ends were then merged based on read names and mapped to MboI restriction fragments using hiclib package (Imakaev et al., 2012). Next, self-circles, dangling ends, PCR duplicates, and genome assembly

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

errors were discarded. Samples of the same cell type were merged. We binned the interaction matrix at different resolutions and corrected it with iterative correction (ICE) for downstream analysis.

Chromatin loops were called with HICCUPS (Durand et al., 2016) for the two different cell types independently. First, we converted the filtered interaction files into juicer format with juicertools. Chromatin loops were called using juicer HICCUPS with bin sizes iterated from 10kb to 25kb by 1kb intervals and parameters "-k VC_SQRT -p 1 -i 3". Only reproducible loops were retained and the highest resolution of the overlapping loops was used.

Topological associated domains (TADs) were identified with Topdom (Shin et al., 2016) at 10K resolution and a 200Kb window size.

Genetic variants concordance analysis

To check for sample mix-ups, we evaluated the genetic variants between WGS genotyping and RNA-seq, ATAC-seq, and ChIP-seq samples. For RNA-seq, ChIP-seq, and ATAC-seq, we called genotypes with GATK (v3.5.0) (McKenna et al., 2010). We performed i) indel-realignment, ii) base score recalibration and iii) joint genotype calling across all samples. Variants with a phred-scaled confidence threshold <10, within ENCODE blacklisted regions of the genome (Amemiya et al., 2019), outside of dbSNP v151, minor allele frequencies (MAF) <25%, and clustered variants were not used. The genotype concordance amongst samples was quantified using both the fraction of concordant genotype calls and the kinship coefficient from KING (v1.9) (Manichaikul et al., 2010), where the two methods give consistent results.

Molecular Data Processing (CMC Cohort)

Genotyping data processing

For the CMC-MPP cohort, genotyping was performed on the Illumina Infinium HumanOmniExpressExome 8 v 1.1b chip (Catalog #: WG-351-2301) using the manufacturer's protocol. For the CMC HBCC cohort, genotyping was performed on one of 3 different Illumina gene chips: HumanHap650Y, Human1M-Duo, and HumanOmni5M-Quad, according to the manufacturer's instructions.

Genotyping QC and imputation proceeded separately by gene chipset as previously described (Hoffman et al., 2019). Briefly, markers with: zero alternate alleles, genotyping call rate ≤ 0.98, Hardy-Weinberg P value < $5x10^{-5}$, and individuals with genotyping call rate < 0.90 were removed by PLINK (Purcell et al., 2007). Then, samples were imputed to HRC (r1.1 2016) (McCarthy et al., 2016). Strands, alleles, position, reference/alternate allele assignments and frequencies were checked; and SNPs that with: A/T & G/C alleles and minor allele frequency (MAF) > 0.4, differing alleles, > 0.2 allele frequency difference between the genotyped samples and the HRC samples, and not in reference panel were removed. Imputation was performed with the Michigan Imputation Server (Das et al., 2016) using Eagle (v2.3) (Loh et al., 2016). All imputation was performed using the GRCh37, the genotype data were subsequently lifted over to GRCh38 for downstream integration with functional genomics assays.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

RNA-seq data processing

RNA-seq data were processed using the Multi-omics RNA-seq pipeline, as described above. The RNA-seq libraries were deeply sequenced and have acceptable values for total read (mean 5.18×10^7, sd $\pm 2.61 \times 10^7$), final read pairs (mean 4.78×10^7, sd $\pm 2.35 \times 10^7$), intergenic rate (mean 5.75%, sd ±2.69%), intronic rate (mean 38.6%, sd ±10.3%), ribosomal RNA rate (mean 0.103%, sd ±0.299%), and RIN (mean 7.41, sd ±0.977).

CAGE-seq data processing

CAGE-seq data of four brain lobes including occipital, frontal, temporal, cerebellum were downloaded from the published result (NCBI Bioproject PRJNA273171) (Yao et al., 2015). We mapped the clean reads to the human reference genome hg38 with STAR (2.5.3a) aligner. The BAM files of the four brain regions were merged for downstream analysis.

eRNA identification

Training set. To define a training set, we first collected the FANTOM5 brain-associated eRNAs that overlapped with our H3K27ac peaks (FANTOM5 eRNAs, FEs). Given that FANTOM5 eRNAs only cover a small fraction of eRNAs, and could have bias, we next determined the CAGE tags at the non-coding OCRs with R bamsignals package (v1.14.0) in a strandspeicifc mode. The analysis yielded 6,810 neuronal and 5,509 non-neuronal bidirectionally expressed enhancers (EEs), as well as 97,827 neuronal and 45,387 non-neuronal not expressed enhancers (NEs). We subsequently compared the epigenomic and transcriptomic profiles at the three sets of enhancers with the bamsignals package (https://bioconductor.org/packages/bamsignals), confirming that the two expressed enhancer groups exhibit different features compared to the not expressed enhancers. We only used the FEs that overlapped corresponding OCRs for downstream analysis. We used FEs and subsampled the same amount of EEs as the positive set. A comparable number of negative sets is subsampled from the NEs.

Testing set. To define a test set, we focused on ATAC-seq peaks. Peaks that overlapped with an annotated exon as well as a 1 kb region up/downstream (Gencode V30) (Harrow et al., 2012), ribosome DNA loci (Wang et al., 2012), and ENCODE blacklisted regions (Amemiya et al., 2019) were filtered. We extend the resulting peak summit to 500bp based on the distribution of FANTOM5 eRNA size (**Figure S1C**), resulting in 168,841 neuronal and 93,139 non-neuronal enhancers.

Prediction model. We collected tag counts with ATAC-seq, ChIP-seq, and strand-specific RNA-seq bam files for the central 200bp and flanking 400bp of both the testing and training sets (**Figure 1D**). Additionally, we annotated the position relative to genes to avoid the effect of premature mRNA on intronic eRNAs. The BAM counts and genomic annotation of the training set were used as the input for a random forest model (Breiman, 2001), with parameters fine-tuned by a 10-fold cross-validation grid search (Kuhn, 2008). The performance of the resulting models was determined by AUROC and AUPC (Sing et al., 2005), both of which generate values of ~0.95 (**Figure 1E and 1F**). We subsequently predict the eRNA from testing sets with the trained models for both cell types.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Quantification. We counted the number of reads overlapped with the eRNAs using featureCounts function in RSubread (v1.6.3) (Liao et al., 2014). The eRNA-gene combined expression matrix (counts per million, CPM) was used for the following analysis.

Differential analysis

We performed differential analysis between neuronal and non-neuronal cells for gene-eRNA of RNA-seq, and peaks of ATACseq and ChIP-seq from the Multi-omics cohort, as well as the differential analysis between SCZ and control gene-eRNA of RNA-seq from the CMC cohort. For RNA-seq analysis only protein-coding genes, lincRNA, and eRNAs were used. A consensus read count-based differential analysis pipeline was used with the following steps:

Read count and transcript/peak filtering. For each assay, transcript/peak read count matrices were used as input. For ATAC-seq, H3K4me3, and H3K27ac ChIP-seq, only peaks that had at least 1 count per million reads (CPM) in > 10% of the samples were retained. For Multi-omics RNA-seq data, we used a relatively low cut-off to account for the modest expression of eRNAs and retained all the transcripts that had at least 0.25 CPM in >10% of the samples. The retained eRNAs that overlapped between neurons and non-neurons were merged and used for downstream analysis. For the CMC data, we contained all the transcripts that had at least 0.1 CPM in >40% of the samples. Read counts were then normalized with the trimmed mean of M-values (TMM) method (Robinson and Oshlack, 2010).

Exploration of covariates and model selection. We performed a PCA on the normalized read count matrix for each assay to identify high-variance components that explained at least 1% of the variance. Correlation tests were performed between the selected PCs and known covariates, and covariates with FDR<0.05 were used for the following steps. To select the final covariates, we first chose the covariates that are known to play a critical role for each assay as “a base model”: for instance, cell types, brain regions, and sex were selected as the base model for the Multi-omics cohort. We then applied an approach based on the Bayesian information criterion (BIC) to select the final covariates (Fromer et al., 2016). We examined the BIC changes in the linear regression model after adding a new covariate, which will be included if it can improve the mean BIC by at least 4. The full list of selected covariates can be found in the table below.

	assay/cohort	base model	BIC selected covariates
Neuron vs Non-Neuron	RNA-Seq	cell	Fraction of gene and eRNA read counts
		brain region	
		sex	
	ATAC-Seq	cell	Fraction of reads with GC 20-39%
		brain region	
		sex	
	H3K4me3	cell	NSC (spp)
		brain region	Fraction of reads in non-blacklisted peaks
		sex	Fraction of reads with GC 40-59%
	H3K27ac	cell	Fraction of reads in non-blacklisted peaks
		brain region	Fraction of reads with GC 80-100%

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

		sex	Fraction of reads in blacklisted peaks vs in all peaks
			Fraction of reads mapped to multiple loci
CMC SCZ vs control	CMC discovery	Dx	Fraction of eRNA reads
		brain region	Fraction of reads with GC 20-39%
		sex	Fraction of intronic reads
		RIN	Fraction of reads mapped to multiple loci
		batch	
		AOD	
	CMC replicate	Dx	Fraction of eRNA reads
		brain region	Fraction of reads with GC 20-39%
		sex	Fraction of UTR reads
		RIN	
		AOD	
	CMC joint	Dx	Fraction of eRNA reads
		brain region	Fraction of reads with GC 20-39%
		sex	Fraction of intronic reads
		RIN	Fraction of reads mapped to multiple loci
		batch	Fraction of UTR reads
		AOD	

Statistical test. With the selected covariates, the normalized read counts were modeled with the voomWithQualityWeights function from the limma package (v.3.38.3) (Ritchie et al., 2015), which utilizes both sample-level and observational-level weights. We subsequently perform the test against the contrast between cell type or disease status using a linear mixed model to account for repeated measurements (i.e. 2 brain regions per individual) in the dream function (Hoffman and Roussos, 2020) of the variancePartition package (Hoffman and Schadt, 2016). Additionally, we determined the level of difference by estimating the proportion of true non-null tests π_1 (Storey and Tibshirani, 2003) with the limma package.

Furthermore, we decomposed variation into known biological and technical factors with VarianceParition (v1.21.2) (Hoffman and Schadt, 2016). The analysis was performed by modeling the $\log_2$ CPM with a linear mixed model and treating each variable as a random effect (Hoffman and Roussos, 2020). Results were summarized in terms of the fraction of total variation explained by each variable for each peak/transcripts.

TF enrichment analysis

Motif selection. 3,059 human TF motifs were downloaded from CIS-BP(v1.02) meta-database (Weirauch et al., 2014). Given that a TF could be represented by multiple motifs, and TFs within the same TF family often share binding motifs, we reduced motifs by merging similar motifs as previously described (Fullard et al., 2018). Briefly, for a given TF, pairwise similarities between motifs were assessed by TomTom (Bailey et al., 2009) with parameters “tomtom -dist kullback -query-pseudo 0.1 -target-pseudo 0.1 -text -min overlap 0 -thresh 1”. Only the motif showing the lowest score was chosen, resulting in 431 motifs representing 798 TFs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ATAC-seq motif enrichment. We used ATAC-seq to examine the motif enrichment between neuronal and non-neuronal cells. All peaks were resized to 500bp centered at the peak summit. For the overlapped peaks between cell types, only the peak with a greater signal value was retained and generated a consensus peak for TF analysis. For eRNA motif analysis, we also used the eRNA overlapped consensus peak for TF analysis instead of eRNA positions. To remove the difference from promoters and non-active enhancers, only eRNA overlapped peaks were used for this analysis. GC content corrected motif enrichment was determined using ChromVar with default parameters (Schep et al., 2017).

DE eRNA enriched TF motifs. For DE neuronal and non-neuronal eRNAs, we first determined the GC content and subsampled the remaining peaks to generate a set of background peaks of matched GC content distribution. We then annotated the motif at both DE eRNAs and background with TFBSTools (v1.20.0) (Tan and Lenhard, 2016). Lastly, we performed a binomial test for the motif enrichment between DE eRNAs and GC matched background.

Identifying eRNA target genes

For every gene, we considered all the eRNAs within a $\pm$500 Kb window centered around the TSS (**Figure 4A**) and fit a 10-fold cross-validation lasso model with the glmnet package (v 2.0.18) (Friedman et al., 2010). glmnet selects lambda to minimize the cross-validation prediction error and then selects lambda within 1 sd of the minimum. The eRNAs with non-zero coefficients were selected as linked eRNAs. For the genes that have only a single eRNAs around the window, we performed a correlation test, and only the $P_{bonferroni}<0.01$ were retained. Given that the link pairs were estimated using only control samples, we imputed the gene expression log fold change using eRNA log fold change of SCZ dysregulation.

Gene set enrichment analysis

To explore the function of a gene set, we collected functional gene sets from MSigDB 7.0 (Liberzon et al., 2011), human brain single-cell markers (Lake et al., 2018), and synaptic gene ontology resource (Koopmans et al., 2019). One-tailed Fisher exact tests were used to test the enrichment and significance.

To examine the genetic enrichment of gene sets, we used MAGMA (v 1.07b) (de Leeuw et al., 2015) with GWAS data (Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium and Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2018; Demontis et al., 2019; Jansen et al., 2019; Karlsson Linnér et al., 2019; Nagel et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2020; Stahl et al., 2019). Briefly, genes were padded by 35kb upstream and 10kb downstream, and the MHC region was removed due to its extensive linkage disequilibrium and complex haplotypes. The European panels from 1000 Genome Project phase 3 were used to estimate the Linkage disequilibrium (LD) (1000 Genomes Project Consortium et al., 2015).

MEGENA gene coexpression network analysis

To identify gene-eRNA co-expressed modules, the selected covariates, except for disease status, were regressed out from the combined gene-expression matrix. We selected the *cis-gene-*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

eRNA-linked transcripts and built co-expression networks using PMFG algorithm and multiscale clustering analysis with MEGENA (v1.3.7) (Song and Zhang, 2015). Only module sizes between 100 to 3000 were selected, resulting in 153 parent-child modules for downstream analysis.

To prioritize SCZ associated modules we tested the significance of (i) SCZ up-regulated transcripts enrichment, (ii) SCZ down-regulated transcripts enrichment, (iii) eQTL overrepresentation of SCZ GWAS signal, (iv) SCZ TWAS significant transcript enrichment. To determine the SCZ up/down regulated transcript enrichment, we generated a gene/eRNA ratio matched background for each module and determined the enrichment with a binomial test. For TWAS enrichment, we do the binomial test and account for both numbers of *cis*-heritable transcripts and gene/eRNA ratios. For eQTL GWAS enrichment, we used a permutation test (Walker et al., 2020). Briefly, the SCZ GWAS P value was assigned to the eSNP LD blocks (r^2>0.8) for every transcript within a module. Then, 1000 permutations were performed, randomly selecting the same number of genes and eRNAs. Significance was determined by the proportion of permuted p values that were larger than the observed P value at expected P value of 0.001. Then the 4 p values were scaled and converted to *Z* scores. We prioritized the modules based on the sum of the 4 *Z* scores. Only modules with an FDR <0.05 in at least one of the four groups and with no parent-child relationship were selected for downstream analysis, yielding 9 SCZ up-regulated modules, 5 SCZ down-regulated modules, and 1 SCZ GWAS eQTL enriched module M26. It is worth noting that M26 is also nominally significant for TWAS transcripts.

To examine the evolutionary conservation of the eRNAs, we first collected human gained enhancers (HGE) (Reilly et al., 2015). We lifted over our eRNA position to hg19 and selected the eRNAs overlapped with HGEs. Additionally, we assigned each eRNA a conservation score as the mean of the corresponding phyloP conservation scores (PhyloP 100-way vertebrate) (Pollard et al., 2006). We selected eRNAs with the top 5% as well as the bottom 5% conservation score as highly/poorly conserved eRNAs (Trevino et al., 2020). One-tailed Fisher exact tests were performed to test the significance of enrichment of the three classes of eRNAs.

We performed key driver analysis for SCZ GWAS significant module M26 using the R KDA package (Zhang and Zhu, 2013). The package defines a background sub-network by looking for neighborhood K-steps (K = 6 in this analysis) away from each node in the target gene list in the network. Then, stemming from each node in this sub-network, it assesses the enrichment in its k-step (k varies from 1 to K) downstream neighborhood for the target gene list.

<u>Module TF enrichment</u>. To identify the driving TF of a module, we first conducted a permutation test to examine motif enrichment. We used the above-described consensus peaks for TF analysis that overlapped with eRNAs to do the analysis.1000 permutations were performed, and eRNA were chosen condition on GC content to have a matched GC content distribution. The significance was determined by the proportion of permutations with a higher motif enrichment. After that, we performed a Spearman correlation test between the TF gene expression and the module eigengene. We prioritized the TF with the highest Spearman correlation coefficient (SCC) with eigengene and permutation TF enrichment p<0.05.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Partitioned heritability analysis

We partitioned heritability for DE eRNAs as well as eSNPs to examine the enrichment of common variants in neuropsychiatric traits with stratified LD score regression (v.1.0.0) (Finucane et al., 2015) from a selection of GWAS studies (Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium and Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2018; Demontis et al., 2019; Jansen et al., 2019; Karlsson Linnér et al., 2019; Nagel et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2020; Stahl et al., 2019). Briefly, with the differential peaks/eRNAs as well as eSNPs, a binary annotation was created by marking all HapMap3 SNPs (International HapMap Consortium, 2003) that fell within the peak or eSNPs and outside the MHC regions. LD scores were calculated for the overlapped SNPs using an LD window of 1cM using 1000 Genomes European Phase LD reference panel (1000 Genomes Project Consortium et al., 2015). The enrichment was determined against the baseline model (Finucane et al., 2015). To enable comparisons of the regression coefficients across traits with a wide range of heritabilities, we chose to normalize it by the per-SNP heritability and named this adjusted metric the “heritability coefficient”. This is not the same as the “enrichment” also outputted by the software, since the heritability coefficient takes the aforementioned baseline into account and the “enrichment” does not.

PRS calculation

For genetic analysis including PRS, eQTL, MESC, FUSION, and FOCUS, only individuals with European ancestry were used. We inferred principal components (PC) of genetic ancestry with GEMTOOLs (Klei et al., 2011). With the first two PCs of the reported European individuals, we determined the center (mean) and standard deviation (sd). Only reported European individuals within 3*sd of the center was selected for the following analysis (**Figure S5D**).

Polygenic scores were constructed for SCZ using effect sizes from the PGC3 schizophrenia GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2020). The summary statistics were processed using PRS-CS software (Ge et al., 2019) to generate weights (posterior SNP effect sizes). Default settings were used for calculating weights using PRS-CS (γ-γ prior=1; parameter b in γ-γ prior=0.5; MCMC iterations=1000; number of burn-in iterations=500; thinning of the Markov chain factor=5). Then, based on the derived weights, individual-level polygenic scores for E-factors were calculated using Plink v2.0 (Chang et al., 2015).

eQTL analysis

We used the MMQTL package to identify *cis*-QTLs for both genes and eRNAs (Zeng et al., 2021). Briefly, Probabilistic Estimation of Expression Residuals (PEER) was used to determine a set of latent covariates to control for unknown biological and technical effects for the two brain regions, independently (Stegle et al., 2012). We use 30 PEER factors for ACC and 35 PEER factors for DLPFC based on the highest percentage of transcripts with eQTLs. The expression matrices were then adjusted for the selected covariates, the first 3 genetic principal components, and PEER factors. We performed a meta-analysis with the normalized expression matrices as well as the SNPs within $\pm$ 1Mb *cis*-window using a linear mixed model to maximize power (Zeng

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

et al., 2021). p values for both GeQTL and EeQTL were corrected together for multiple testing using Storey and Tibshirani FDR correction (Storey and Tibshirani, 2003).

The proportion of null-hypotheses (π_0) was estimated by looking up the corresponding p values of the significant GTEx eQTL pairs (FDR<0.05) in our result with the qvalue package (Storey and Tibshirani, 2003). The non-null-hypothesis π_1=1-π_0 was subsequently determined. Additionally, we calculated the Spearman correlation coefficient between the effect sizes of our GeQTL and GTEx brain eQTL with all significant eQTLs (FDR<0.05). We performed a similar analysis to brain hQTL (Sun et al., 2016) and caQTL (Bryois et al., 2018). Given that both analyses are done with hg19, we lifted over eRNA to hg19 and found the overlapped peaks of the eRNA positions. With the overlapped significant peak-SNP pairs, we calculated the π_1 using EeQTL p values for corresponding pairs.

To explore the functional enrichment of gene and eRNA eQTL, we performed enrichment analysis using Genomic Regulatory Elements and Gwas Overlap algoRithm (GREGOR) (Schmidt et al., 2015) with our ChromHMM results. Briefly, we grouped genes and eRNAs based on the differential expression status between neurons and non-neurons. The corresponding eSNPs were extended to all SNPs in high linkage disequilibrium (r^2>0.7) and determined enrichment.

MESC analysis

We used the mediated expression score regression (MESC) package (Yao et al., 2020) to estimate SCZ heritability mediated by the *cis* genetic component of expression levels of gene, eRNA, or combined respectively. Based on the principle that expression mediated effect size introduces a linear relationship between the eQTL effect sizes and disease effect sizes, MESC determined the mediated effect through modeling GWAS summary statistics, LD scores, and eQTL effect sizes (Yao et al., 2020). 1000 Genomes European Phase reference panel (1000 Genomes Project Consortium et al., 2015) were used to compute LD scores. We retained only the SNPs that were from HapMap 3 (International HapMap Consortium, 2003) for this analysis. GCTA was used to estimate the *cis*-heritability ($\pm$ 500kb) of each transcript for two brain regions independently (Yang et al., 2011). Then we used the LASSO model from PLINK to estimate the eQTL size for each transcript. We performed a meta-analysis to combine the two brain regions to determine the expression score. Lastly, we estimated the SCZ heritability mediated by the expression of eRNA, gene, and combined transcripts.

TWAS

The FUSION package was used for the gene-eRNA TWAS analysis (Gusev et al., 2016) for the two brain regions independently. For each brain region, only *cis*-heritable transcripts (GCTA nominal p<0.05 and *cis*-heritability>0) were used for the following analysis. We built expression models using best *cis*-eQTL, Elastic-net regression, and LASSO regression with five-fold cross-validation (Barbeira et al., 2018). The model with the best R^2 was selected. Then we performed the association analysis with PGC3 SCZ GWAS summary statistics (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al., 2020). MHC regions were excluded for this analysis. The resulting p values of genes and eRNAs are together Bonferroni corrected. $P_{bonferroni}$<0.05 were selected for downstream analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We performed colocalization analysis using the bayesian estimator COLOC (v4.0.4) (Giambartolomei et al., 2014, 2018) has been incorporated in the TWAS/FUSION pipeline. COLOC examine the posterior probability of five hypotheses: H0, no eQTL and no GWAS association; H1 and H2, associated to either eQTL or GWAS but not both; H3, eQTL and GWAS association but independent; H4, association. The analysis yields corresponding posterior probability PP0-PP4.

FOCUS

To prioritize the potential casual transcripts for TWAS analysis, we performed statistical fine-mapping using FOCUS (Mancuso et al., 2019). FOCUS models the correlation structure induced by LD and prediction weights in TWAS, and controls for certain pleiotropic effects. To account for the genes that are filtered by low *cis*-heritability, we collected gene prediction results from GTEx expression panels. For each gene, the model with the best accuracy was included. With gene/eRNA prediction results from two brain regions and the GTEx models, we performed FOCUS fine-mapping and got the 90% credible transcript sets.

Supplementary Tables

Table S1 identified the eRNA list
Table S2 Differential analysis results between neuronal and non-neuronal cells of different assays
Table S3 Differential expression summary statistics between SCZ and control
Table S4 gene-enhancer link summary
Table S5 Statistics for Significant eQTLs, top SNP per transcript
Table S6 TWAS fine-mapped transcripts and the eRNA-linked genes
Table S7 MEGENA Module membership

References

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE blacklist: identification of problematic regions of the genome. Sci. Rep. *9*, 9354.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Baggelaar, M.P., Maccarrone, M., and van der Stelt, M. (2018). 2-Arachidonoylglycerol: A signaling lipid with manifold actions in the brain. Prog. Lipid Res. *71*, 1–17.

Bahl, E., Koomar, T., and Michaelson, J.J. (2017). cerebroViz: an R package for anatomical visualization of spatiotemporal brain data. Bioinformatics *33*, 762–763.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202-8.

Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. *9*, 1825.

Bhattaram, P., Penzo-Méndez, A., Sock, E., Colmenares, C., Kaneko, K.J., Vassilev, A., Depamphilis, M.L., Wegner, M., and Lefebvre, V. (2010). Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. Nat. Commun. *1*, 9.

Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium and Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2018). Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell *173*, 1705-1715.e16.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Breiman, L. (2001). Random Forests. Springer Science and Business Media LLC.

Bryois, J., Garrett, M.E., Song, L., Safi, A., Giusti-Rodriguez, P., Johnson, G.D., Shieh, A.W., Buil, A., Fullard, J.F., Roussos, P., et al. (2018). Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. Nat. Commun. *9*, 3121.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1-21.29.9.

Catarino, R.R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev. *32*, 202–223.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

Chen, H., Li, C., Peng, X., Zhou, Z., Weinstein, J.N., Cancer Genome Atlas Research Network, and Liang, H. (2018). A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. Cell *173*, 386-399.e12.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat. Genet. *46*, 1311–1320.

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.

Delaneau, O., Zazhytska, M., Borel, C., Giannuzzi, G., Rey, G., Howald, C., Kumar, S., Ongen, H., Popadin, K., Marbach, D., et al. (2019). Chromatin three-dimensional interactions mediate genetic effects on gene expression. Science *364*.

DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.-D., Williams, C., Reich, M., Winckler, W., and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

optimization. Bioinformatics *28*, 1530–1532.

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. *51*, 63–75.

Dietz, A.G., Goldman, S.A., and Nedergaard, M. (2020). Glial cells in schizophrenia: a unified hypothesis. Lancet Psychiatry *7*, 272–281.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dong, X., Liao, Z., Gritsch, D., Hadzhiev, Y., Bai, Y., Locascio, J.J., Guennewig, B., Liu, G., Blauwendraat, C., Wang, T., et al. (2018). Enhancers active in dopamine neurons are a primary link between genetic variation and neuropsychiatric disease. Nat. Neurosci. *21*, 1482–1492.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. *3*, 95–98.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. *12*, 2478–2492.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. *33*, 1–22.

Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. *19*, 1442–1453.

Fudenberg, G., and Imakaev, M. (2017). FISH-ing for captured contacts: towards reconciling FISH and 3C. Nat. Methods *14*, 673–678.

Fullard, J.F., Hauberg, M.E., Bendl, J., Egervari, G., Cirnaru, M.-D., Reach, S.M., Motl, J., Ehrlich, M.E., Hurd, Y.L., and Roussos, P. (2018). An atlas of chromatin accessibility in the adult human brain. Genome Res. *28*, 1243–1252.

Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*.

van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods *12*, 1061–1063.

Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat. Commun. *10*, 1776.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.

Giambartolomei, C., Zhenli Liu, J., Zhang, W., Hauberg, M., Shi, H., Boocock, J., Pickrell, J., Jaffe, A.E., CommonMind Consortium, Pasaniuc, B., et al. (2018). A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics *34*, 2538–2545.

Gill, K.M., and Grace, A.A. (2014). The role of α5 GABAA receptor agonists in the treatment of cognitive deficits in schizophrenia. Curr. Pharm. Des. *20*, 5069–5076.

Girdhar, K., Hoffman, G.E., Jiang, Y., Brown, L., Kundakovic, M., Hauberg, M.E., Francoeur, N.J., Wang, Y.-C., Shah, H., Kavanagh, D.H., et al. (2018). Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. Nat. Neurosci. *21*, 1126–1136.

Golson, M.L., and Kaestner, K.H. (2016). Fox transcription factors: from development to disease. Development *143*, 4558–4570.

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444.

Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., et al. (2015). Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell *162*, 1051–1065.

GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

Guillemot, F. (2007). Spatial and temporal specification of neural fates by transcription factor codes. Development *134*, 3771–3780.

Guo, J., Higginbotham, H., Li, J., Nichols, J., Hirt, J., Ghukasyan, V., and Anton, E.S. (2015). Developmental disruptions underlying brain abnormalities in ciliopathies. Nat. Commun. *6*, 7857.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.

Gusev, A., Mancuso, N., Won, H., Kousi, M., Finucane, H.K., Reshef, Y., Song, L., Safi, A., Schizophrenia Working Group of the Psychiatric Genomics Consortium, McCarroll, S., et al. (2018). Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat. Genet. *50*, 538–548.

Gu, B., Swigut, T., Spencley, A., Bauer, M.R., Chung, M., Meyer, T., and Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science *359*, 1050–1055.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Hauberg, M.E., Fullard, J.F., Zhu, L., Cohain, A.T., Giambartolomei, C., Misir, R., Reach, S., Johnson, J.S., Wang, M., Mattheisen, M., et al. (2019). Differential activity of transcribed enhancers in the prefrontal cortex of 537 cases with schizophrenia and controls. Mol. Psychiatry *24*, 1685–1695.

Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hoffman, G.E., and Roussos, P. (2020). dream: Powerful differential expression analysis for repeated measures designs. Bioinformatics.

Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics *17*, 483.

Hoffman, G.E., Hartley, B.J., Flaherty, E., Ladran, I., Gochman, P., Ruderfer, D.M., Stahl, E.A., Rapoport, J., Sklar, P., and Brennand, K.J. (2017). Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. Nat. Commun. *8*, 2225.

Hoffman, G.E., Bendl, J., Voloudakis, G., Montgomery, K.S., Sloofman, L., Wang, Y.-C., Shah, H.R., Hauberg, M.E., Johnson, J.S., Girdhar, K., et al. (2019). CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. Sci. Data *6*, 180.

Hoser, M., Potzner, M.R., Koch, J.M.C., Bösl, M.R., Wegner, M., and Sock, E. (2008). Sox12 deletion in the mouse reveals nonreciprocal redundancy with the related Sox4 and Sox11 transcription factors. Mol. Cell. Biol. *28*, 4675–4687.

Hou, T.Y., and Kraus, W.L. Spirits in the material world: enhancer rnas in transcriptional regulation. Trends Biochem. Sci. *46*, 138–153.

Huckins, L.M., Dobbyn, A., Ruderfer, D.M., Hoffman, G., Wang, W., Pardiñas, A.F., Rajagopal, V.M., Als, T.D., T Nguyen, H., Girdhar, K., et al. (2019). Publisher Correction: Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. Nat. Genet. *51*, 1068.

Hunt, G.J., Freytag, S., Bahlo, M., and Gagnon-Bartsch, J.A. (2019). dtangle: accurate and robust cell type deconvolution. Bioinformatics *35*, 2093–2099.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Imbrici, P., Camerino, D.C., and Tricarico, D. (2013). Major channels involved in neuropsychiatric disorders and therapeutic perspectives. Front. Genet. *4*, 76.

International HapMap Consortium (2003). The international hapmap project. Nature *426*, 789–796.

Jaffe, A.E., Straub, R.E., Shin, J.H., Tao, R., Gao, Y., Collado-Torres, L., Kam-Thong, T., Xi, H.S., Quan, J., Chen, Q., et al. (2018). Developmental and genetic regulation of the human cortex

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

transcriptome illuminate schizophrenia pathogenesis. Nat. Neurosci. *21*, 1117–1125.

Jansen, P.R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A.R., de Leeuw, C.A., Benjamins, J.S., Muñoz-Manchado, A.B., Nagel, M., et al. (2019). Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat. Genet. *51*, 394–403.

Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S.F.W., Wedow, R., Fontana, M.A., Lebreton, M., Tino, S.P., Abdellaoui, A., Hammerschlag, A.R., et al. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat. Genet. *51*, 245–257.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Klei, L., Kent, B.P., Melhem, N., Devlin, B., and Roeder, K. (2011). GemTools: A fast and efficient approach to estimating genetic ancestry. ArXiv.

Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., et al. (2019). SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. Neuron *103*, 217-234.e4.

Krijger, P.H.L., and de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. Nat. Rev. Mol. Cell Biol. *17*, 771–782.

Kuhn, M. (2008). Building Predictive Models in *R* Using the caret Package. J. Stat. Softw. *28*.

Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., et al. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat. Biotechnol. *36*, 70–80.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Liem, K.F., He, M., Ocbina, P.J.R., and Anderson, K.V. (2009). Mouse Kif7/Costal2 is a cilia-associated protein that regulates Sonic hedgehog signaling. Proc Natl Acad Sci USA *106*, 13377–13382.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., et al. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science *362*.

Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. *48*, 1443–1448.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187.

Mahar, M., and Cavalli, V. (2018). Intrinsic mechanisms of neuronal axon regeneration. Nat. Rev. Neurosci. *19*, 323–337.

Mancuso, N., Freund, M.K., Johnson, R., Shi, H., Kichaev, G., Gusev, A., and Pasaniuc, B. (2019). Probabilistic fine-mapping of transcriptome-wide association studies. Nat. Genet. *51*, 675–682.

Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.

Marques, T.R., Ashok, A.H., Angelescu, I., Borgan, F., Myers, J., Lingford-Hughes, A., Nutt, D.J., Veronese, M., Turkheimer, F.E., and Howes, O.D. (2020). GABA-A receptor differences in schizophrenia: a positron emission tomography study using [11C]Ro154513. Mol. Psychiatry.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., and Furlong, E.E.M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev. *32*, 42–57.

Murakawa, Y., Yoshihara, M., Kawaji, H., Nishikawa, M., Zayed, H., Suzuki, H., Fantom Consortium, and Hayashizaki, Y. (2016). Enhanced Identification of Transcriptional Enhancers Provides Mechanistic Insights into Diseases. Trends Genet. *32*, 76–88.

Nagel, M., Watanabe, K., Stringer, S., Posthuma, D., and van der Sluis, S. (2018). Item-level analyses reveal genetic heterogeneity in neuroticism. Nat. Commun. *9*, 905.

Ogasawara, D., Deng, H., Viader, A., Baggelaar, M.P., Breman, A., den Dulk, H., van den Nieuwendijk, A.M.C.H., Soethoudt, M., van der Wel, T., Zhou, J., et al. (2016). Rapid and profound rewiring of brain lipid signaling networks by acute diacylglycerol lipase inhibition. Proc Natl Acad Sci USA *113*, 26–33.

Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., NISC Comparative Sequencing Program, et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

variants. Proc Natl Acad Sci USA *110*, 17921–17926.

Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat. Rev. Genet. *14*, 288–295.

Poirier, R., Cheval, H., Mailhes, C., Garel, S., Charnay, P., Davis, S., and Laroche, S. (2008). Distinct functions of egr gene family members in cognitive processes. Front. Neurosci. *2*, 47–55.

Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel, A., Pedersen, J.S., Bejerano, G., Baertsch, R., et al. (2006). Forces shaping the fastest evolving regions in the human genome. PLoS Genet. *2*, e168.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

Raivich, G., and Behrens, A. (2006). Role of the AP-1 transcription factor c-Jun in developing, adult and injured brain. Prog. Neurobiol. *78*, 347–363.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Reilly, S.K., Yin, J., Ayoub, A.E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P., and Noonan, J.P. (2015). Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science *347*, 1155–1159.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Rothschild, G., and Basu, U. (2017). Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. Trends Genet. *33*, 143–154.

Roussos, P., Katsel, P., Davis, K.L., Siever, L.J., and Haroutunian, V. (2012). A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples. Arch. Gen. Psychiatry *69*, 1205–1213.

Rudolph, U., and Möhler, H. (2014). GABAA receptor subtypes: Therapeutic potential in Down syndrome, affective disorders, schizophrenia, and autism. Annu. Rev. Pharmacol. Toxicol. *54*, 483–507.

Sartorelli, V., and Lauberth, S.M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. Nat. Struct. Mol. Biol. *27*, 521–528.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

975–978.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke, S., Walters, J.T., and O'Donovan, M.C. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. MedRxiv.

Schmidt, E.M., Zhang, J., Zhou, W., Chen, J., Mohlke, K.L., Chen, Y.E., and Willer, C.J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinformatics *31*, 2601–2606.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. *20*, 437–455.

Schrode, N., Ho, S.-M., Yamamuro, K., Dobbyn, A., Huckins, L., Matos, M.R., Cheng, E., Deans, P.J.M., Flaherty, E., Barretto, N., et al. (2019). Synergistic effects of common schizophrenia risk variants. Nat. Genet. *51*, 1475–1485.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.

Shin, H., Shi, Y., Dai, C., Tjong, H., Gong, K., Alber, F., and Zhou, X.J. (2016). TopDom: an efficient and deterministic method for identifying topological domains in genomes. Nucleic Acids Res. *44*, e70.

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCR: visualizing classifier performance in R. Bioinformatics *21*, 3940–3941.

Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar, H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado, A.B., et al. (2018). Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. *50*, 825–833.

Song, W.-M., and Zhang, B. (2015). Multiscale Embedded Gene Co-expression Network Analysis. PLoS Comput. Biol. *11*, e1004574.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. *51*, 793–803.

Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. *7*, 500–507.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc Natl Acad Sci USA *100*, 9440–9445.

Sullivan, P.F., and Geschwind, D.H. (2019). Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. Cell *177*, 162–183.

Sun, W., Poschmann, J., Cruz-Herrera Del Rosario, R., Parikshak, N.N., Hajan, H.S., Kumar, V., Ramasamy, R., Belgard, T.G., Elanggovan, B., Wong, C.C.Y., et al. (2016). Histone Acetylome-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

wide Association Study of Autism Spectrum Disorder. Cell *167*, 1385-1397.e11.

Tan, G., and Lenhard, B. (2016). TFBSTools: an R/bioconductor package for transcription factor binding site analysis. Bioinformatics *32*, 1555–1556.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tippens, N.D., Liang, J., Leung, A.K.-Y., Wierbowski, S.D., Ozer, A., Booth, J.G., Lis, J.T., and Yu, H. (2020). Transcription imparts architecture, function and logic to enhancer units. Nat. Genet. *52*, 1067–1075.

Trevino, A.E., Sinnott-Armstrong, N., Andersen, J., Yoon, S.-J., Huber, N., Pritchard, J.K., Chang, H.Y., Greenleaf, W.J., and Pașca, S.P. (2020). Chromatin accessibility dynamics in a model of human forebrain development. Science *367*.

Varki, A., Geschwind, D.H., and Eichler, E.E. (2008). Explaining human uniqueness: genome interactions with environment, behaviour and culture. Nat. Rev. Genet. *9*, 749–763.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet. *101*, 5–22.

Wafford, K.A., Macaulay, A.J., Fradley, R., O'Meara, G.F., Reynolds, D.S., and Rosahl, T.W. (2004). Differentiating the role of gamma-aminobutyric acid type A (GABAA) receptor subtypes. Biochem. Soc. Trans. *32*, 553–556.

Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. *51*, 592–599.

Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., de la Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2020). Genetic control of expression and splicing in developing human brain informs disease mechanisms. Cell *181*, 745.

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al. (2018a). Comprehensive functional genomic resource and integrative model for the human brain. Science *362*.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.

Wang, M., Beckmann, N.D., Roussos, P., Wang, E., Zhou, X., Wang, Q., Ming, C., Neff, R., Ma, W., Fullard, J.F., et al. (2018b). The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. Sci. Data *5*, 180185.

Wang, Q., Chen, R., Cheng, F., Wei, Q., Ji, Y., Yang, H., Zhong, X., Tao, R., Wen, Z., Sutcliffe, J.S., et al. (2019). A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. Nat. Neurosci. *22*, 691–699.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.14.443421; this version posted May 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

transcription factor sequence specificity. Cell *158*, 1431–1443.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Won, H., Huang, J., Opland, C.K., Hartl, C.L., and Geschwind, D.H. (2019). Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. Nat. Commun. *10*, 2396.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

Yao, D.W., O'Connor, L.J., Price, A.L., and Gusev, A. (2020). Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. *52*, 626–633.

Yao, P., Lin, P., Gokoolparsadh, A., Assareh, A., Thang, M.W.C., and Voineagu, I. (2015). Coexpression networks identify brain region-specific enhancer RNAs in the human brain. Nat. Neurosci. *18*, 1168–1174.

Yap, E.-L., and Greenberg, M.E. (2018). Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. Neuron *100*, 330–348.

Yu, M., and Ren, B. (2017). The Three-Dimensional Organization of Mammalian Genomes. Annu. Rev. Cell Dev. Biol. *33*, 265–289.

Zeng, B., Bendl, J., Kosoy, R., Fullard, J.F., Hoffman, G.E., and Roussos, P. (2021). Trans-ethnic eQTL meta-analysis of human brain reveals regulatory architecture and candidate causal variants for brain-related traits. MedRxiv.

Zhang, B., and Zhu, J. (2013). Identification of Key Causal Regulators in Gene Networks. In Proceedings of the World Congress on Engineering, (iaeng.org), p.

Zhang, W., Voloudakis, G., Rajagopal, V.M., Readhead, B., Dudley, J.T., Schadt, E.E., Björkegren, J.L.M., Kim, Y., Fullard, J.F., Hoffman, G.E., et al. (2019). Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. Nat. Commun. *10*, 3834.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.