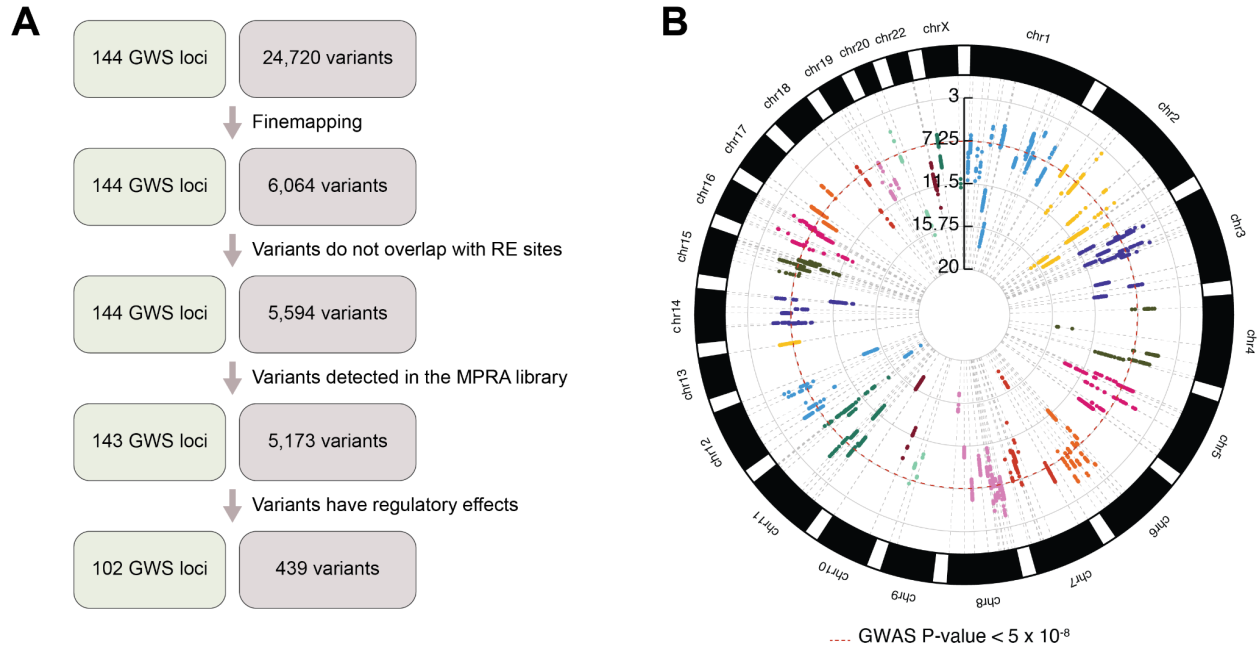# Supplemental information

# Systematic investigation of allelic regulatory

activity of schizophrenia-associated

common variants

Jessica C. McAfee, Sool Lee, Jiseok Lee, Jessica L. Bell, Oleh Krupa, Jessica Davis, Kimberly Insigne, Marielle L. Bond, Nanxiang Zhao, Alan P. Boyle, Douglas H. Phanstiel, Michael I. Love, Jason L. Stein, W. Brad Ruzicka, Jose Davila-Velderrain, Sriram Kosuri, and Hyejung Won

Supplementary Figures

A

144 GWS loci	24,720 variants

↓ Finemapping

144 GWS loci	6,064 variants

↓ Variants do not overlap with RE sites

144 GWS loci	5,594 variants

↓ Variants detected in the MPRA library

143 GWS loci	5,173 variants

↓ Variants have regulatory effects

| 102 GWS loci	439 variants

B

- - - - GWAS P-value < 5 x 10⁻⁸

Figure S1. MPRA prioritizes schizophrenia risk variants, related to Figure 1. A. A schematic of refining the number of causal variants associated with schizophrenia. In total, 24,720 variants are located in 144 GWS loci associated with schizophrenia at $P<1\times10^{-5}$. Computational finemapping identified 6,064 credible variants that can explain the causal configuration of 144 GWS loci. Some of the credible variants were discarded because they overlap with restriction enzyme (RE) sites used for cloning of AAV-MPRA libraries. Some of the variants were missed during the cloning steps, resulting in 5,173 variants to be tested via MPRA. Finally, we identified 439 variants covering 102 GWS loci that show gene regulatory effects. **B.** A circular Manhattan plot displaying the finemapped schizophrenia GWAS variants with their respective P-values. Red dotted line depicts the GWAS P-value threshold of $P<5\times10^{-8}$.
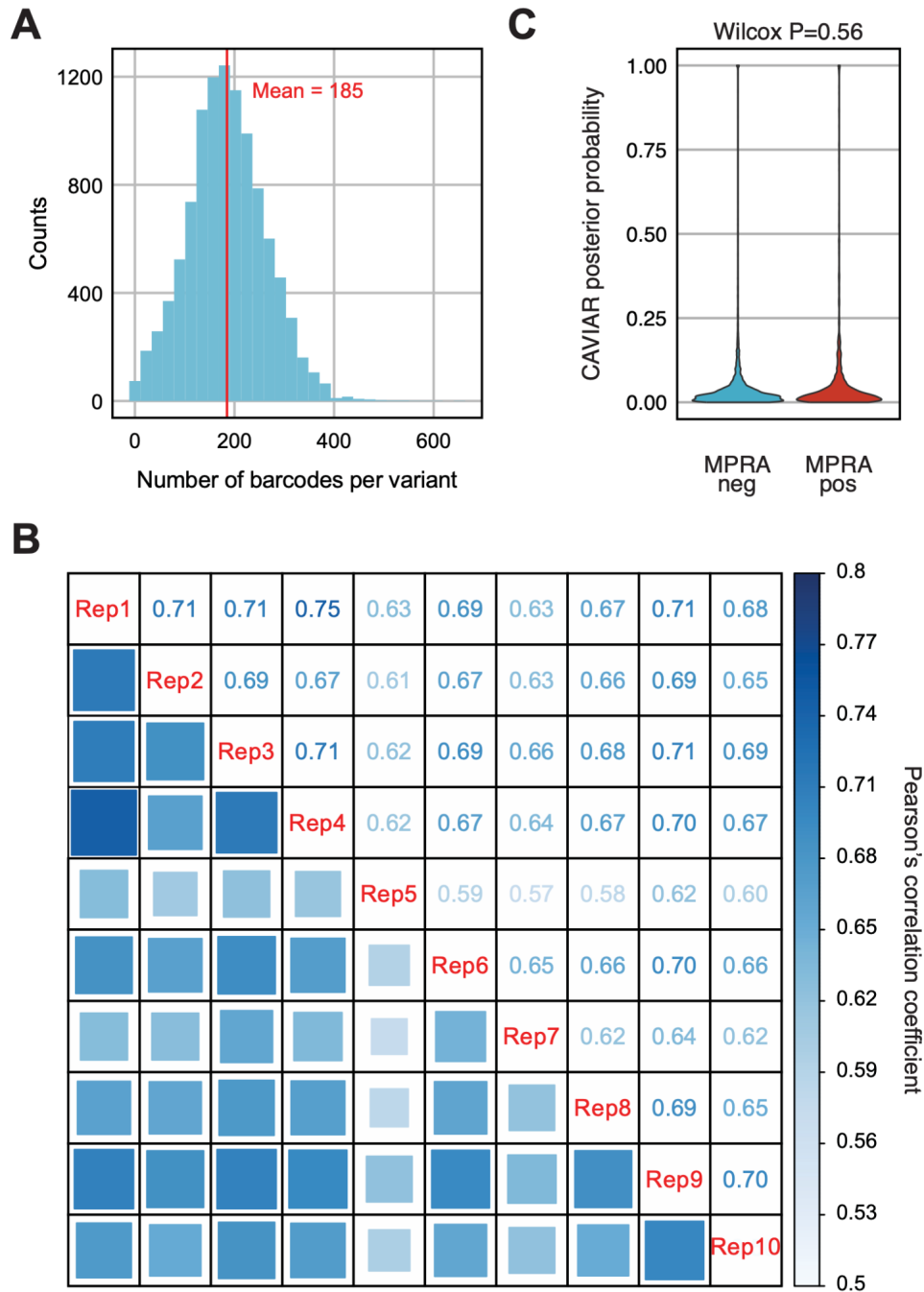
Figure S2. Quality control and reproducibility, related to Figure 1. A. The distribution of the number of barcodes assigned to each variant. **B.** Reproducibility across 10 biological replicates measured by Pearson's correlation coefficients. **C.** Finemap posterior probabilities predicted by CAVIAR do not differ between MPRA-positive (MPRA-pos) and MPRA-negative (MPRA-neg) variants in an independent MPRA dataset[15]. P-values calculated by two-sided Wilcoxon rank sum test.
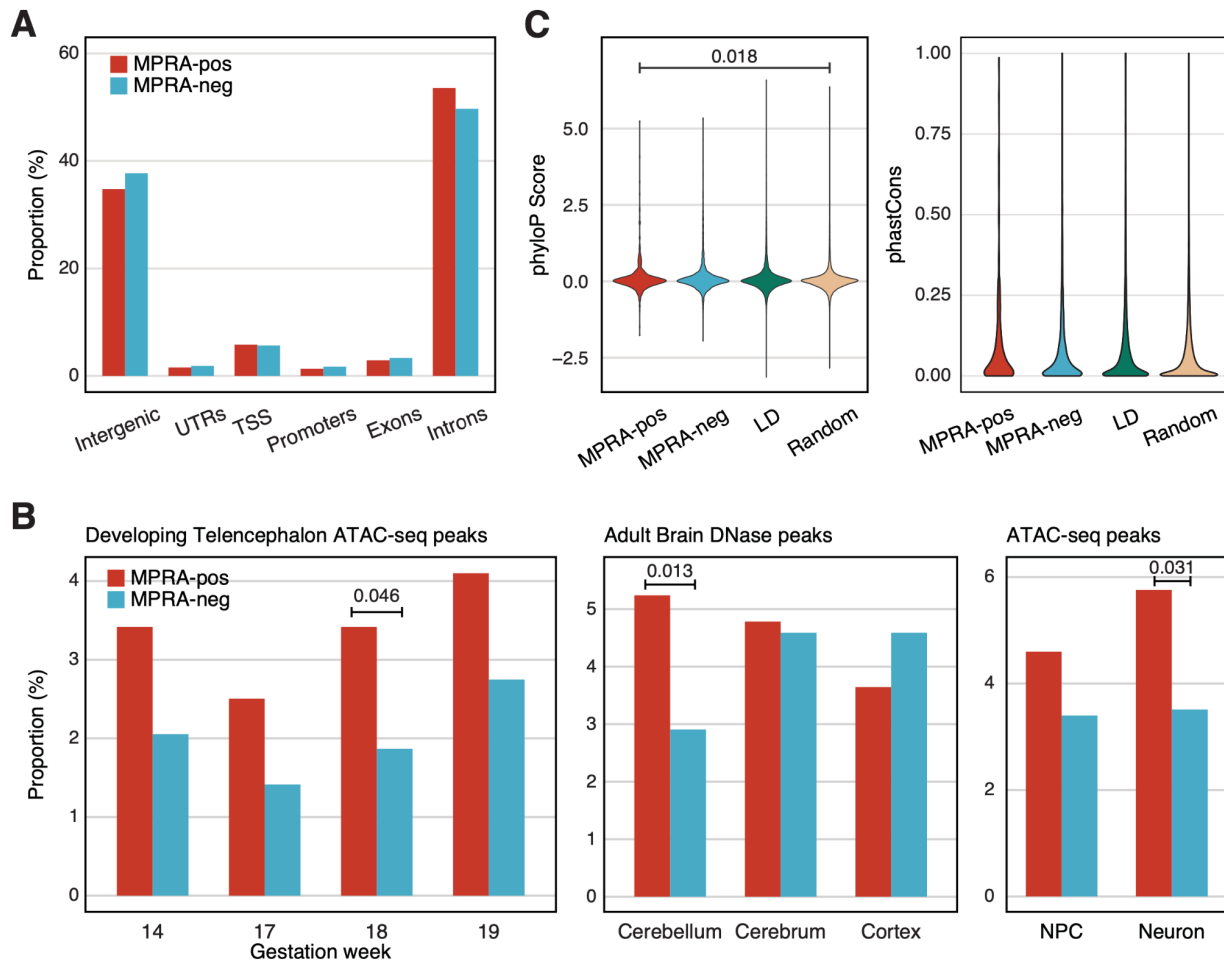
Figure S3. Genomic and epigenomic properties of MPRA-positive variants, related to Figure 2. A. Genomic annotation of MPRA-positive and -negative variants. UTRs, untranslated regions; TSS, transcription start sites. **B.** The proportion of epigenetic overlap of MPRA-positive (MPRA-pos) and MPRA-negative (MPRA-neg) variants to the developing telencephalon ATAC-seq peaks (left), adult brain DNase peaks (middle), and ATAC-peaks from human-induced pluripotent stem cell (hiPSC)-derived neural progenitor cells (NPC) and differentiated neurons (right). P values were calculated by using a two-sided Fisher's exact test **C.** Evolutionary conservation (left, phyloP scores; right, phastCons scores) of MPRA-positive, MPRA-negative, LD, and random SNPs. P values were calculated by using a two-sided Wilcoxon rank sum test.
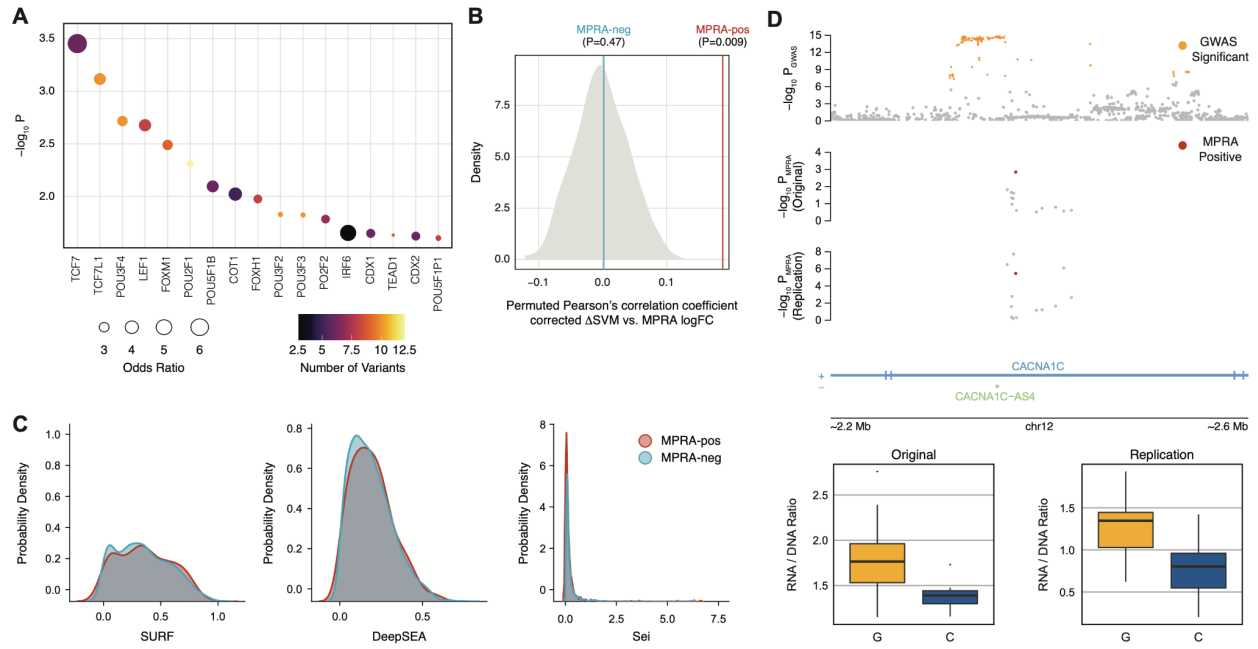
Figure S4. Epigenetic properties that can distinguish MPRA-positive from MPRA-negative variants, related to Figure 2. A. TFs whose motifs are predicted to be altered by MPRA-positive variants. TF enrichment was calculated by comparing TF binding motifs between MPRA-positive variants and random SNPs. Each dot is color-coded based on the number of variants that are predicted to alter TF binding motifs. **B.** Pearson's correlation coefficient between TF-mediated regulation (corrected ΔSVM) and allelic regulatory activity (MPRA logFC) is significantly higher for MPRA-positive variants than permuted distribution; P-values were calculated by permutation **C**. Kernel density estimation (KDE) plots for *SURF*, *DeepSEA*, and *Sei* scores of MPRA-positive and -negative variants. **D.** The MPRA-positive variant (rs11062170, highlighted in red) in the *CACNA1C* locus was replicated in an independent experiment with the identical direction of the effect.
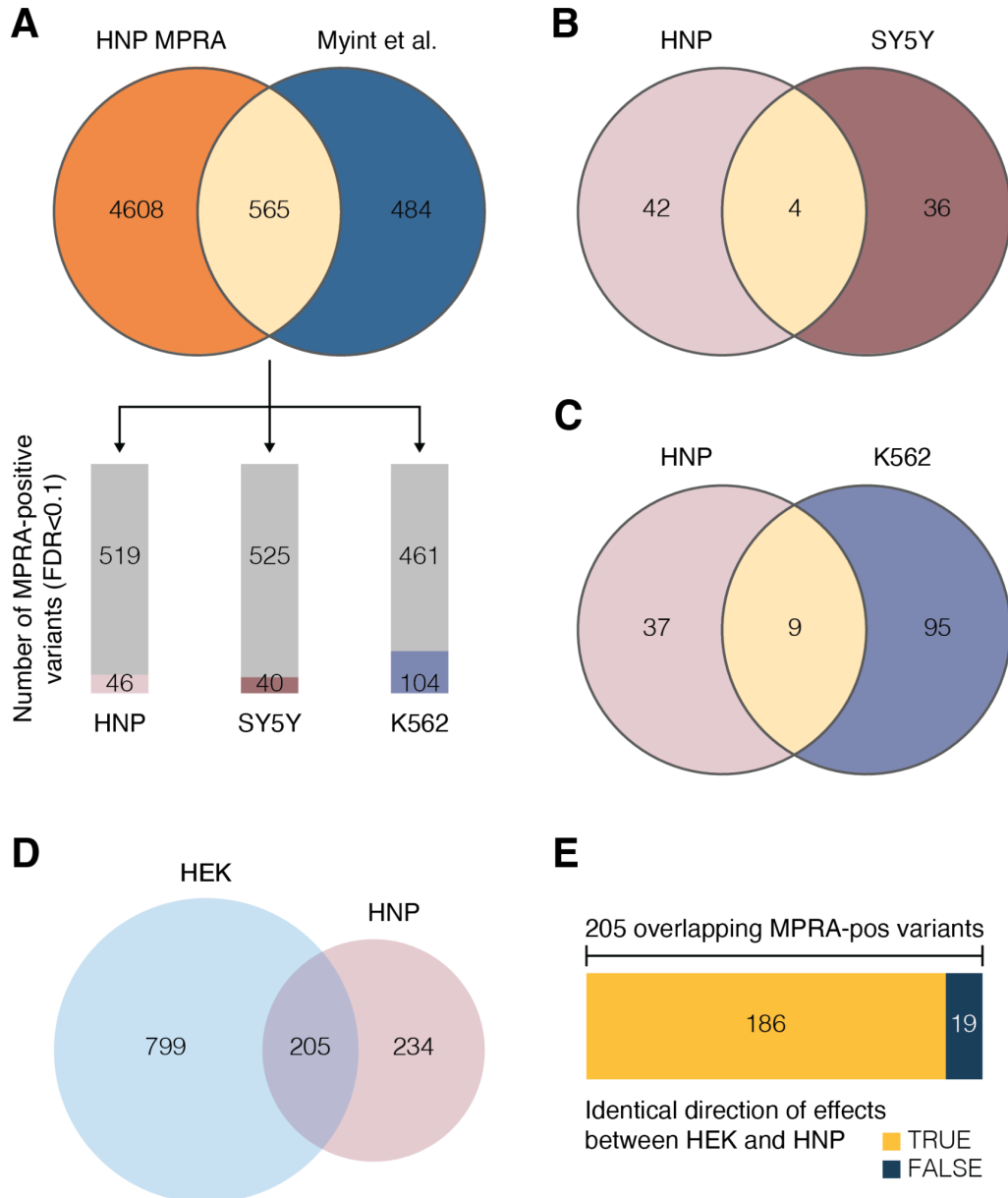
Figure S5. Comparison of MPRA-positive variants defined in different cell types, related to STAR Methods. A. We compared our MPRA results obtained from HNP (HNP MPRA) with the previously reported MPRA results from SY5Y and K562 (Myint et al.)[32]. Top, 565 variants were tested in both studies. Bottom, the number of MPRA-positive variants in HNPs, SY5Y, and K562 at FDR<0.1 among 565 variants tested in both studies. **B.** The overlap of MPRA-positive variants in HNP and SY5Y. **C.** The overlap of MPRA-positive variants in HNP and K562. **D.** The overlap of MPRA-positive variants in HNP and HEK293 cells. **E.** The proportion of MPRA-positive variants that show identical direction of effects between HEK293 cells and HNPs.
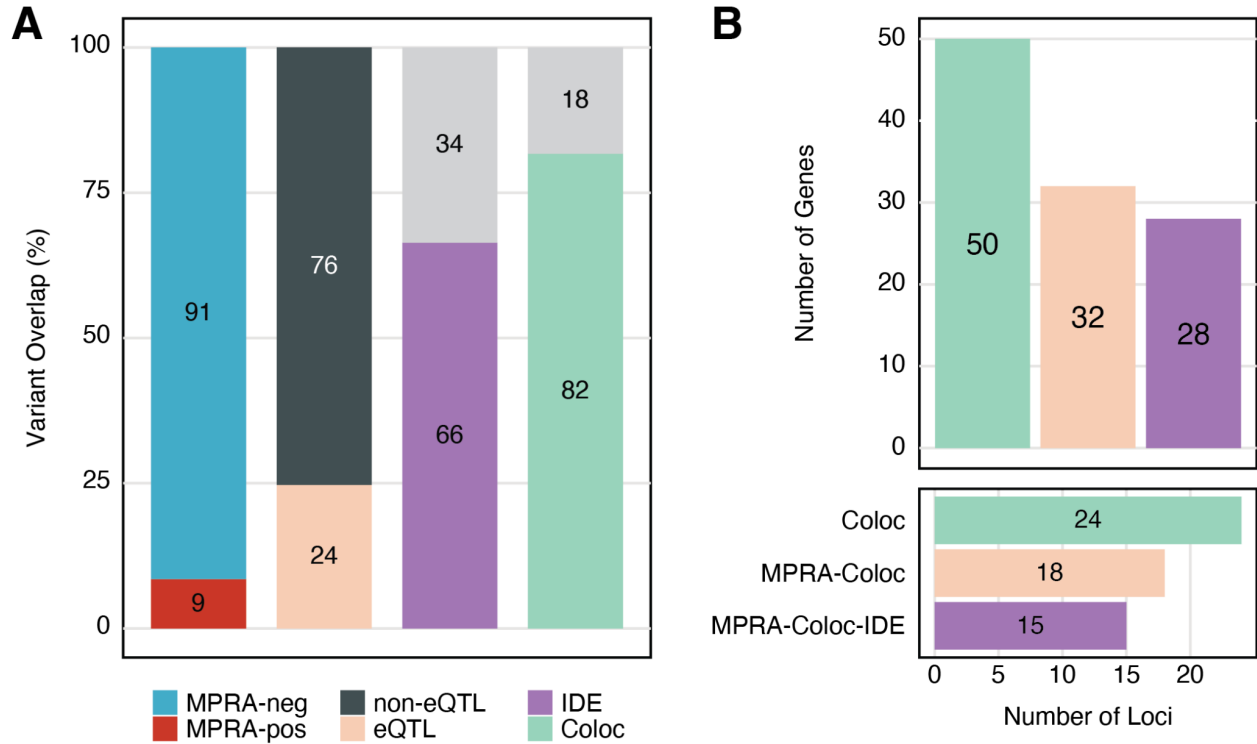
Figure S6. Comparison of MPRA results with developing brain eQTLs, related to Figure 3.
A. 9% of the variants tested in our MPRA were found to have allelic regulatory effects. 24% of MPRA-positive variants overlapped with developing brain eQTLs. Within the 24% overlap with eQTLs, 66% of MPRA-positive variants are identical in expression direction (IDE) to the overlapping eQTL variants. Within that 66%, 82% of IDE variant-gene pairs were detected from the colocalization analysis between eQTLs and schizophrenia GWAS (Coloc). **B.** 50 schizophrenia GWS loci colocalize with eQTLs, providing 24 schizophrenia-associated eGenes (Coloc). 32 out of these 50 loci contain at least one MPRA-positive variant and are mapped to 18 eGenes (MPRA-Coloc). 28 of MPRA-Coloc loci contain variants that have the identical direction of effects between MPRA and eQTLs and are mapped to 15 eGenes (MPRA-Coloc-IDE).
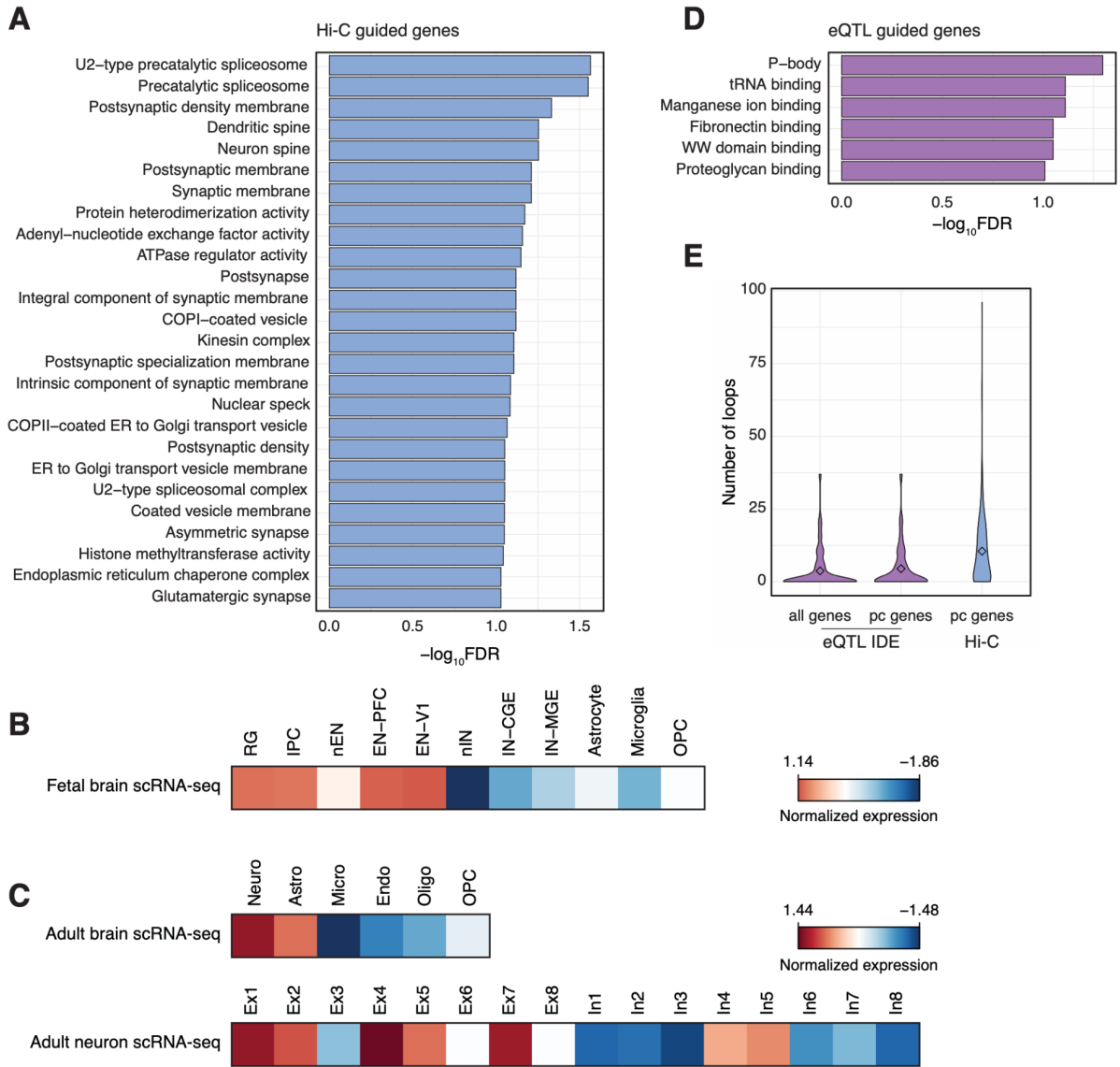
Figure S7. Gene ontology and cell type enrichment of MPRA$_{\text{Hi-C}}$ and MPRA$_{\text{eQTL-IDE}}$ genes, related to Figure 4. A. Top gene ontology terms for MPRA$_{\text{Hi-C}}$ genes. **B.** Average expression of MPRA$_{\text{Hi-C}}$ genes in different cell types of the fetal brain. RG, radial glia; IPC, intermediate progenitor cells; nEN, neonatal excitatory neuron; EN-PFC, excitatory neuron in the prefrontal cortex; EN-V1, excitatory neuron in the visual cortex; nIN, neonatal inhibitory neuron; IN-CGE, inhibitory neuron in the caudal ganglionic eminence; IN-MGE, inhibitory neuron in the medial ganglionic eminence; OPC, oligodendrocyte precursor cell. **C.** Average expression of MPRA$_{\text{Hi-C}}$ genes in different cell types of the adult brain. Neuro, neuron; Astro, astrocyte; Micro, microglia; Endo, endothelial cell; Oligo, oligodendrocyte; Ex1, L2/3 cortical projecting neuron; Ex2, L3/4

granule neuron; Ex3, L4 granule neuron; Ex4, L4 and L3/5/6 subcortical projecting neuron; Ex5, L4 and L5/6 subcortical projecting neuron; Ex6, L6 neuron; Ex7, L5/6 corticothalamic projecting neuron; Ex8, L6 corticothalamic projecting neuron; In1, VIP+/Reln+/NDNF+ interneuron; In2, VIP+/Reln–/NDNF– interneuron; In3, VIP+/Reln+/NDNF– interneuron; In4, VIP–/Reln+/NDNF+ interneuron; In5, CCK+/nNOS+/Calbindin+ interneuron; In6, Parvalbumin+/CRHBP interneuron; In7, Somatostatin+/Calbindin+/NPY+ interneuron; In8, Somatostatin+/nNOS+ interneuron. **D.** Top gene ontology terms for MPRA$_{eQTL-IDE}$ genes. **E.** Number of promoter-anchored loops for MPRA$_{eQTL-IDE}$ genes, MPRA$_{eQTL-IDE}$ protein coding (pc) genes, and MPRA$_{Hi-C}$ genes. Promoter-anchored loops from adult neurons were used to measure regulatory complexity.