The American Journal of Human Genetics, Volume *107*

Supplemental Data

Regional Variation of Splicing QTLs in Human Brain

Yida Zhang, Harry Taegyun Yang, Kathryn Kadash-Edmondson, Yang Pan, Zhicheng Pan, Beverly L. Davidson, and Yi Xing

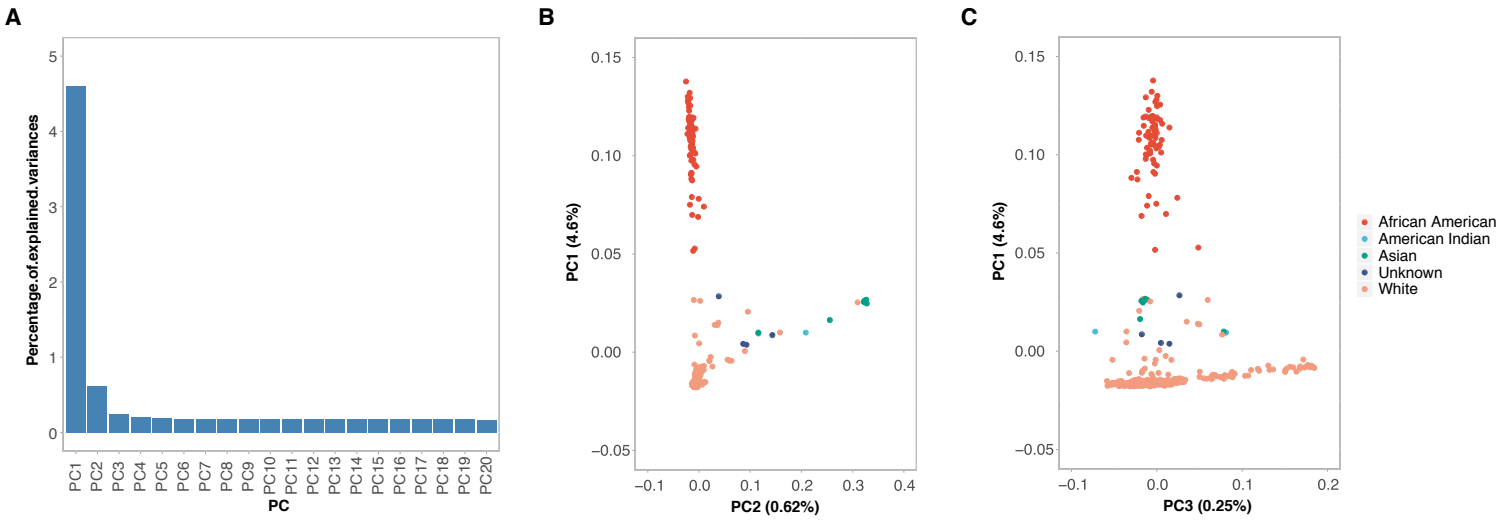


Figure S1. Principal component analysis (PCA) of genotype data of 635 GTEx DNA samples
(A) Plot showing the percentage of explained variances in genotype data by the top 20 principal components (PCs). (B) PCA of the 645 GTEx DNA samples with self-reported ethnicity (PC1 vs. PC2). (C) PCA of the 645 GTEx DNA samples with self-reported ethnicity (PC1 vs. PC3).

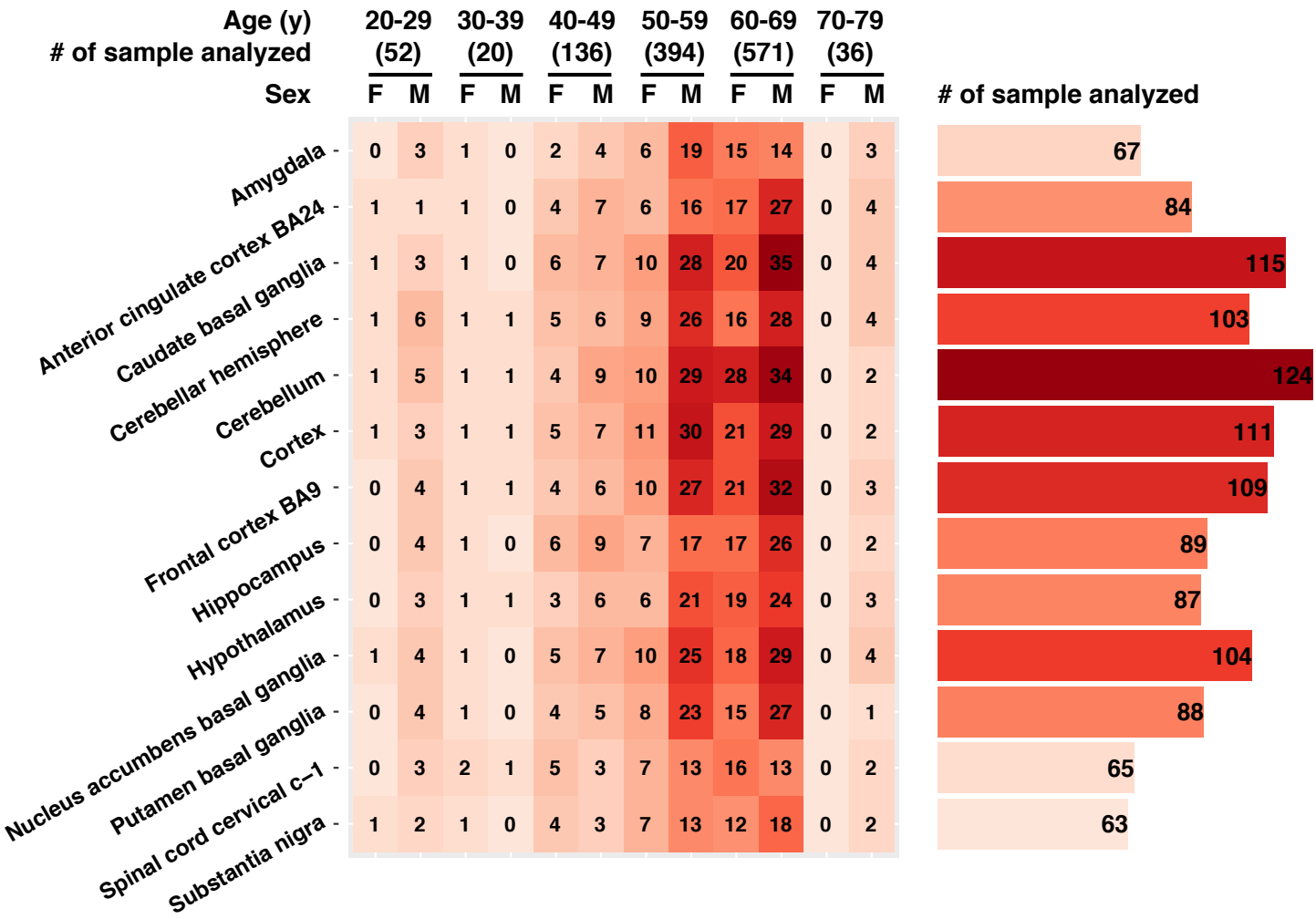


Figure S2. Distribution of GTEx Brain Tissue Samples by Age, Sex, and Brain Region
Heat map showing number of tissue samples with available genotype information for each combination of age, sex, and brain region. Bar plots on right show the total number of samples with available genotype information in each brain region.

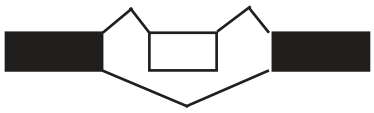
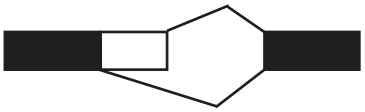
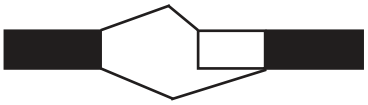
Event Types		Events after filtering
Skipped exon		10,665
Alternative 5’ splice site		1,443
Alternative 3’ splice site		2,434

Figure S3. Numbers of Events for Three Types of Alternative Splicing Patterns
Numbers represent events after filtering. Schematic diagrams of the three types of alternative splicing patterns are shown.

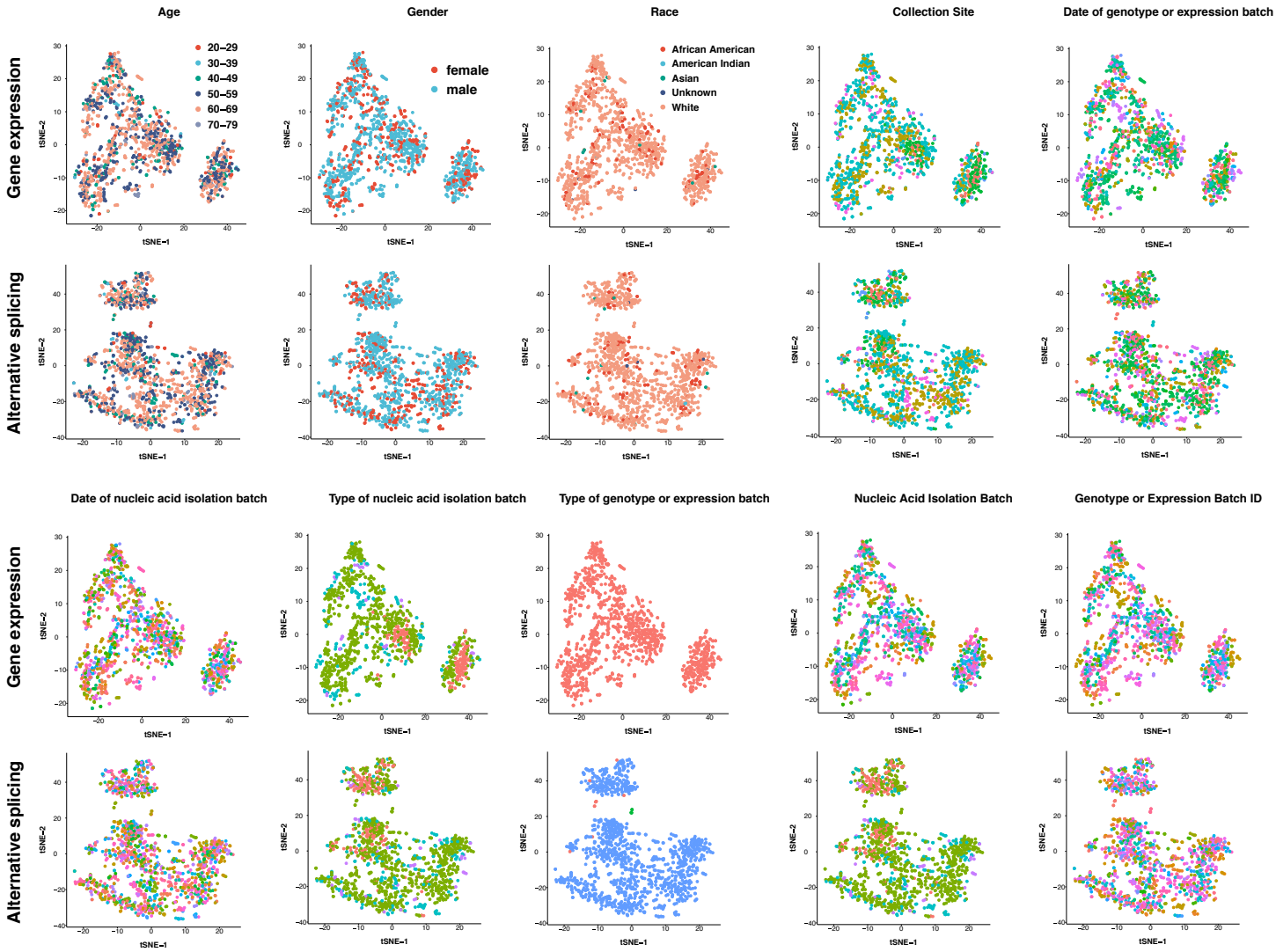


Figure S4. t-SNE Clustering of Brain Samples

t-SNE clustering of all brain tissue samples based on gene expression (upper) and alternative splicing (lower). Samples are color-coded by age, sex, race, and selected potential batch effects.

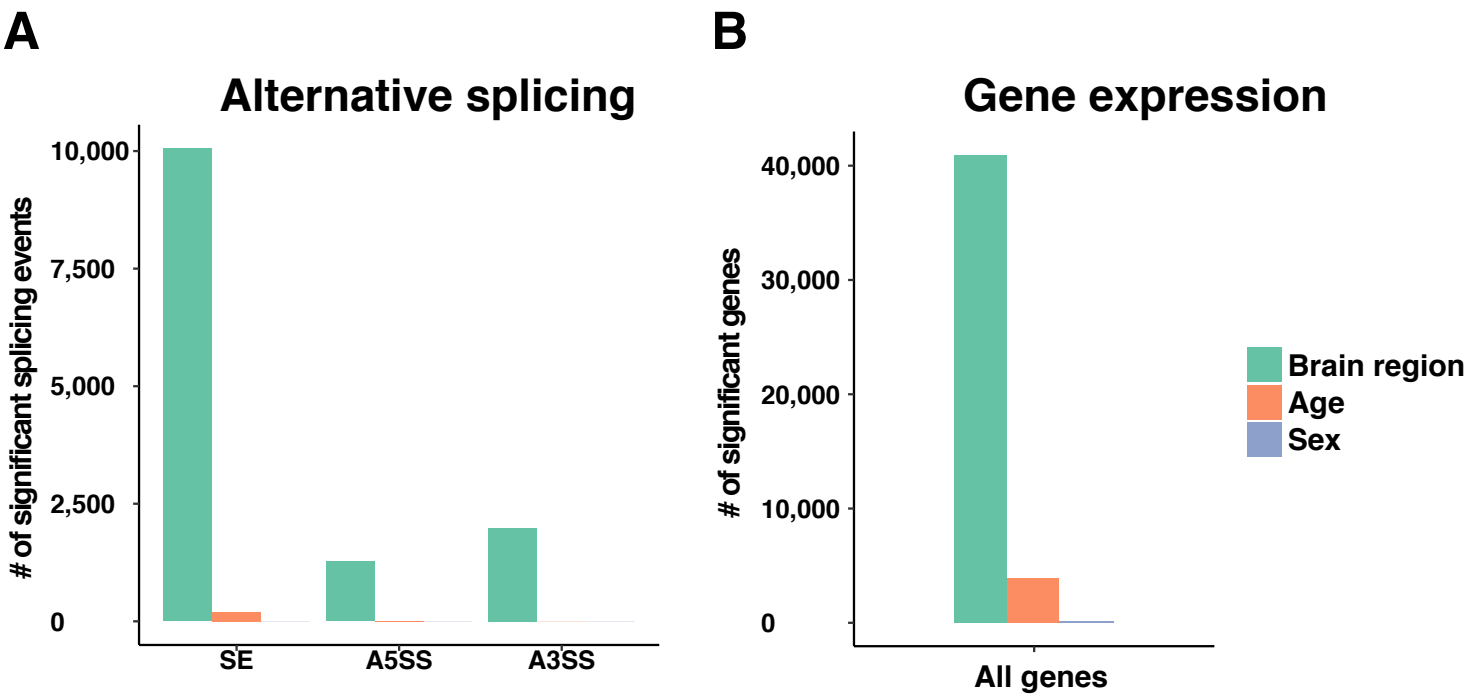


Figure S5. Association of Alternative Splicing and Gene Expression Events with Age, Sex, and Brain Region

Bar plots showing numbers of (A) alternative splicing events (including SE, A5SS, and A3SS events) and (B) expressed genes, with significant differences across age groups, sexes, or brain regions.

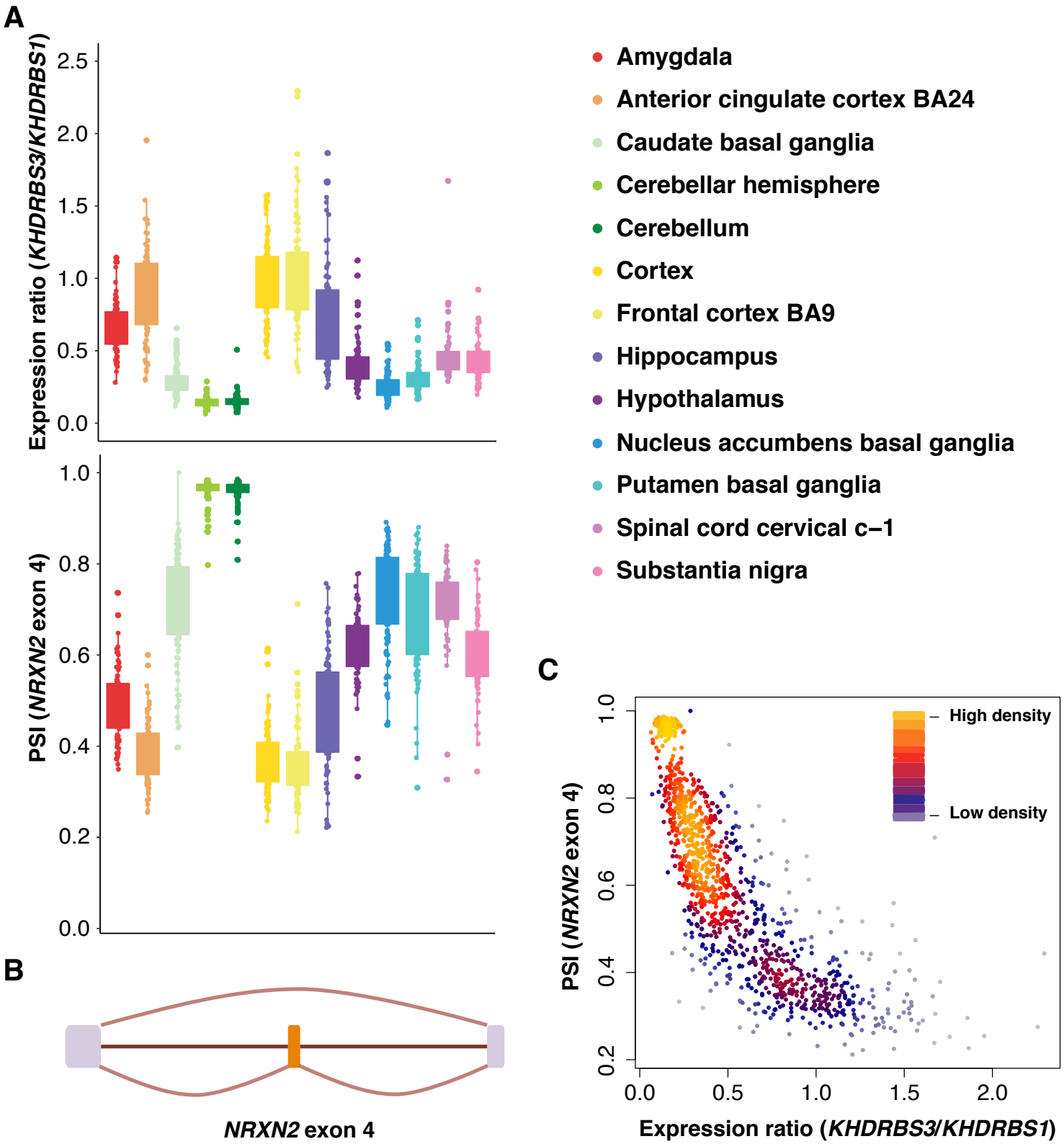


Figure S6. Regulation of Brain Region-Specific Splicing of *NRXN2* Exon 4 by KHDRBS3 and KHDRBS1
(A) (Upper panel) Box plot showing the gene expression ratio between *KHDRBS3* and *KHDRBS1* (*KHDRBS3/KHDRBS1*) in each brain region. (Lower panel) Box plot showing the PSI value (exon inclusion level) of *NRXN2* exon 4 in each brain region. (B) Structure of the alternative splicing event. (C) Scatter plot showing the negative correlation between the gene expression ratio (*KHDRBS3/KHDRBS1*) and the PSI value (exon inclusion level) of *NRXN2* exon 4.

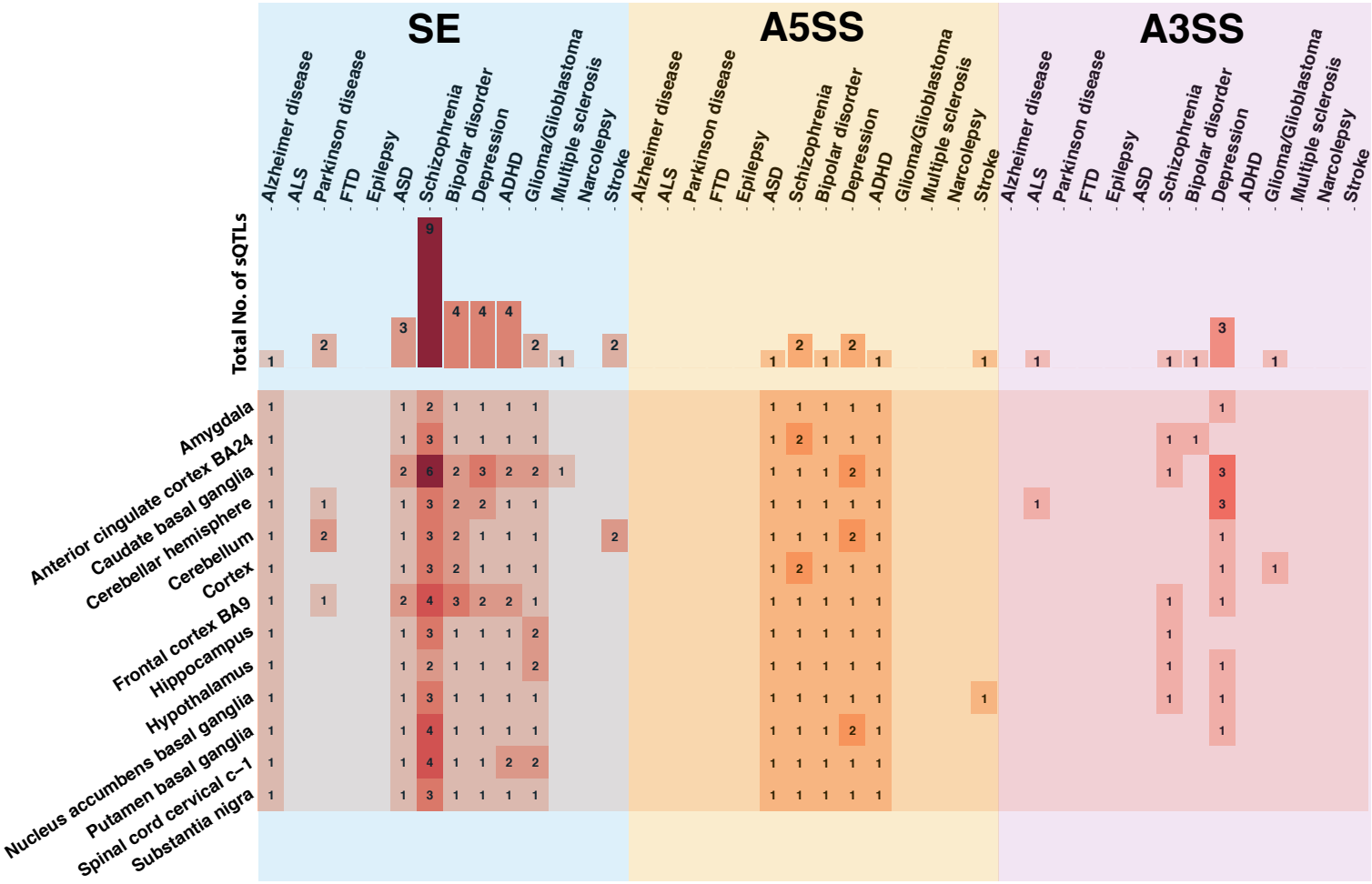


Figure S7. The number of disease sQTLs associated with each neurological disorder in each brain region based on GWAS SNPs reaching genome-wide significance

Heat map showing the number of disease sQTLs (including SE, A5SS, and A3SS events) associated with each neurological disorder in each brain region based on GWAS SNPs reaching genome-wide significance (*P* value ≤ 5×10^{-8}). Each row represents one brain region. Each column represents one neurological disorder. Bar plot above heat map shows the total number of unique sQTLs associated with each neurological disorder.

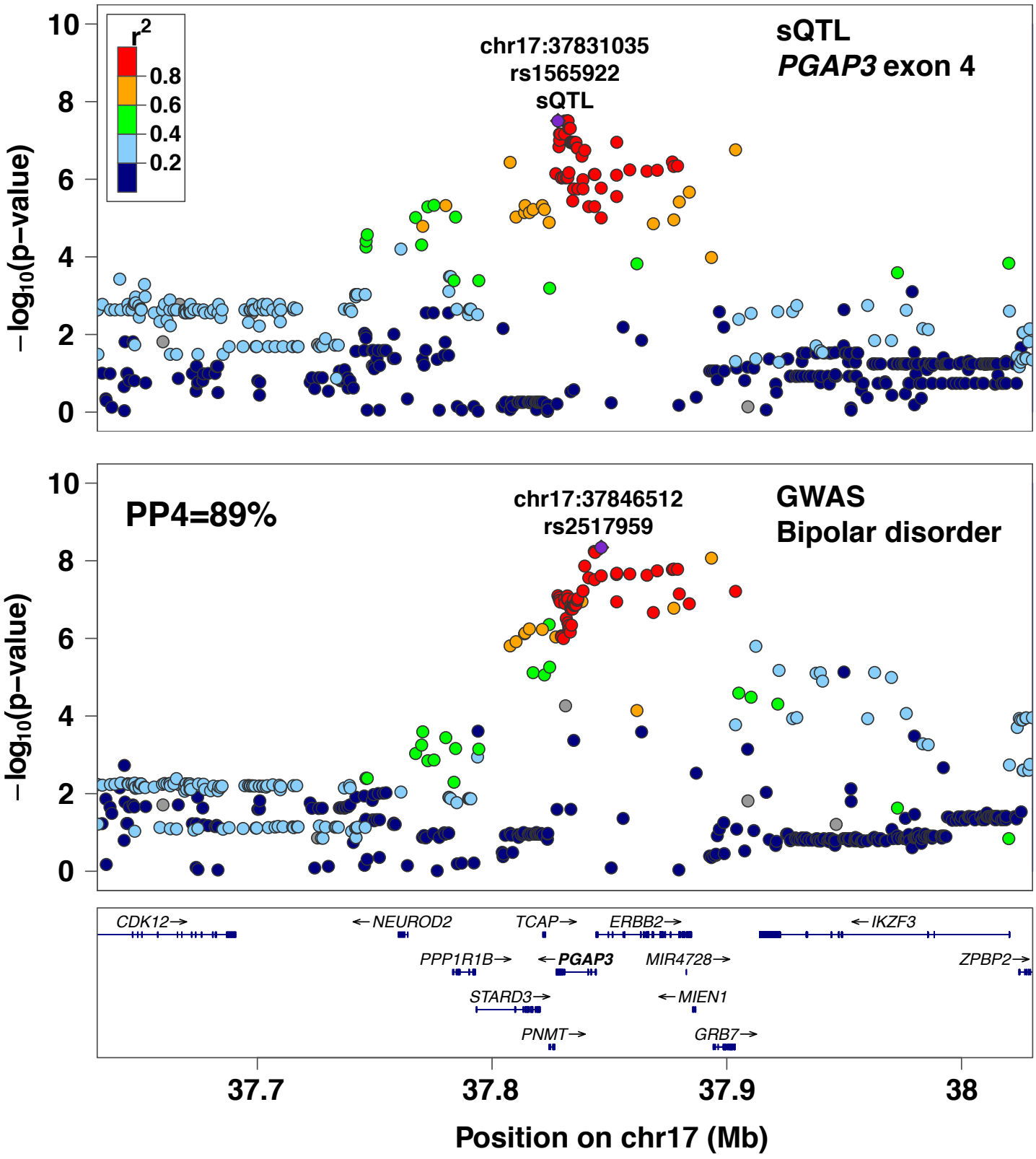


Figure S8. Colocalization between sQTL and GWAS signals at the *PGAP3* locus
Association landscapes for colocalizing associations for the sQTL of *PGAP3* exon 4 in the cortex (upper) and one GWAS trait (bipolar disorder, GWAS *P* value = 5×10^{-9}, lower). Each dot represents one SNP, color-coded according to its LD with the best hit (in purple). Y-axis shows the significance ($-\log_{10}$(*P* value)) of association between each SNP and the sQTL exon in the cortex (upper) and the significance ($-\log_{10}$(*P* value)) of the GWAS association (lower). Genes in the UCSC Genome Browser (https://genome.ucsc.edu/) are shown below the association plots.

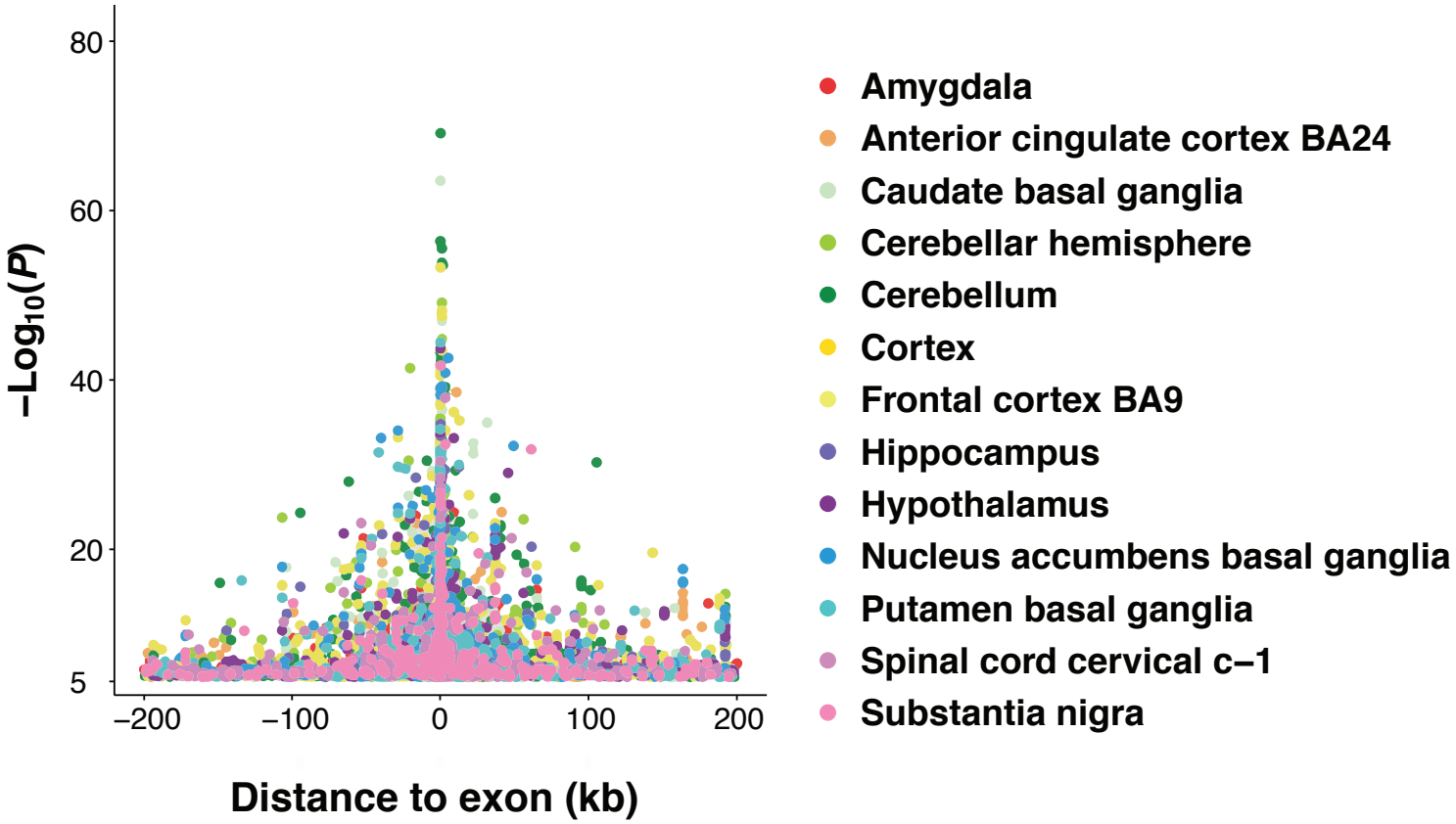


Figure S9. Positional Distribution and Significance of sQTL SNPs
Significance (-$\log_{10}$(*P* value)) of sQTL SNPs in all 13 brain regions as a function of the distance of the sQTL SNP to the nearest alternative splicing event (sQTL exon). Each dot represents one significant sQTL SNP in one brain region. Dot color corresponds to the brain region where the sQTL exon and SNP show a significant association.

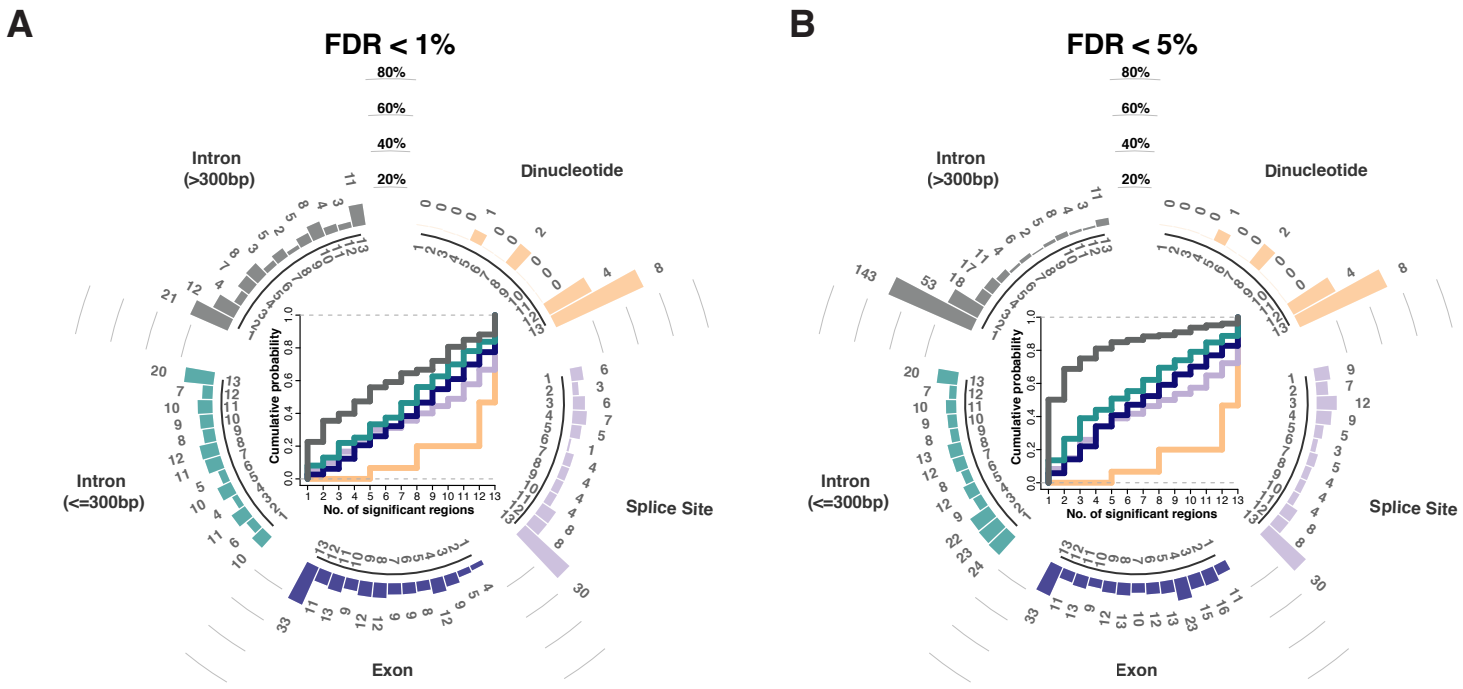


Figure S10. Relationship between SNP Position and Brain Region Specificity of sQTL Exons (SE Events)

Bar plots (outside) and cumulative distribution functions (CDFs) (inside) showing the relationship between SNP position and brain region specificity of sQTLs (SE events). The sQTLs are grouped based on the position of their significant SNP (e.g., dinucleotide, splice site, etc.). For each group, the bar plot shows the histogram of the percentage of sQTLs that are significant in a given number of brain regions. Each bar is labeled above with the number of significant sQTLs in the given number of brain regions. Significant sQTLs were defined with different significance cutoffs (left: FDR < 1%; right: FDR < 5%).

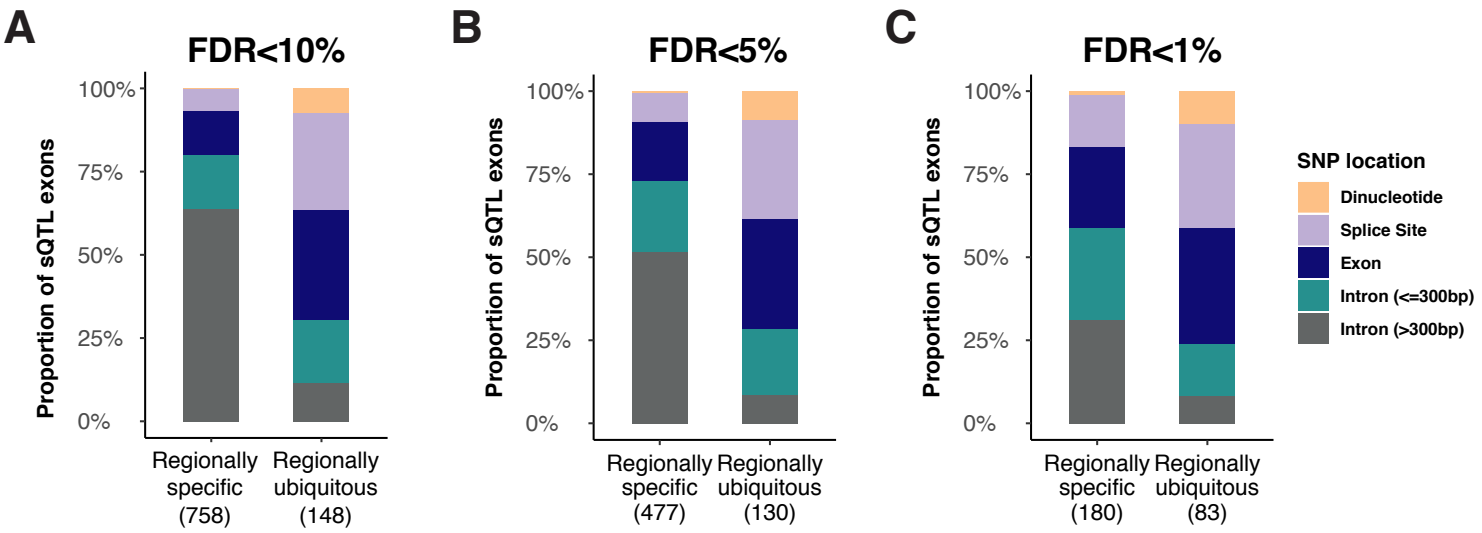
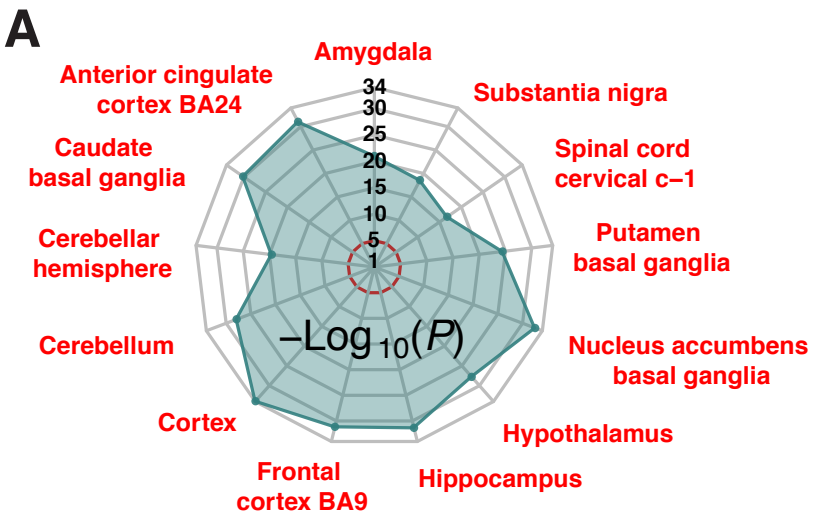


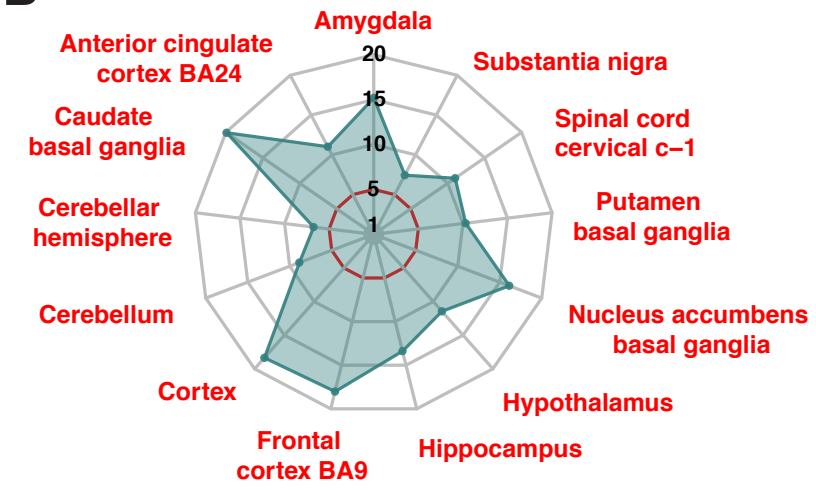
Figure S11. Positional Distributions of SNPs for Regionally Specific and Regionally Ubiquitous sQTLs

Stacked bar plots showing proportions of sQTLs (grouped based on significant SNP positions) for regionally specific and regionally ubiquitous sQTLs. Significant sQTLs are defined with different significance cutoffs (A: FDR < 10%; B: FDR < 5%; C: FDR < 1%).



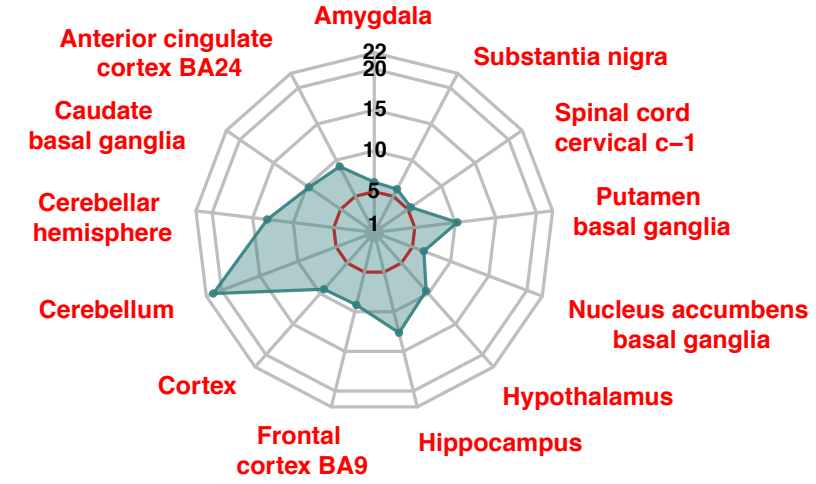
Gene: *C8orf59*
Exon: 2
SNP: rs67573812

Position: 5' splice site (dinucleotide)
GWAS association: NA



Gene: *FLOT1*
Exon: 5
SNP: rs1059612

Position: Intron (≤300bp)
GWAS association: schizophrenia, ADHD, unipolar depression, autism spectrum disorder, bipolar disorder



Gene: *SLC39A13*
Exon: 5
SNP: rs2293576

Position: Exon
GWAS association: Alzheimer's disease

Figure S12. Examples of Regionally Ubiquitous sQTLs
(A), (B), (C) Radar plots showing significance (-$\log_{10}$(*P* value)) of three regionally ubiquitous sQTLs (A: rs67573812 and *C8orf59* exon 2; B: rs1059612 and *FLOT1* exon 5; C: rs2293576 and *SLC39A13* exon 5) in each brain region.

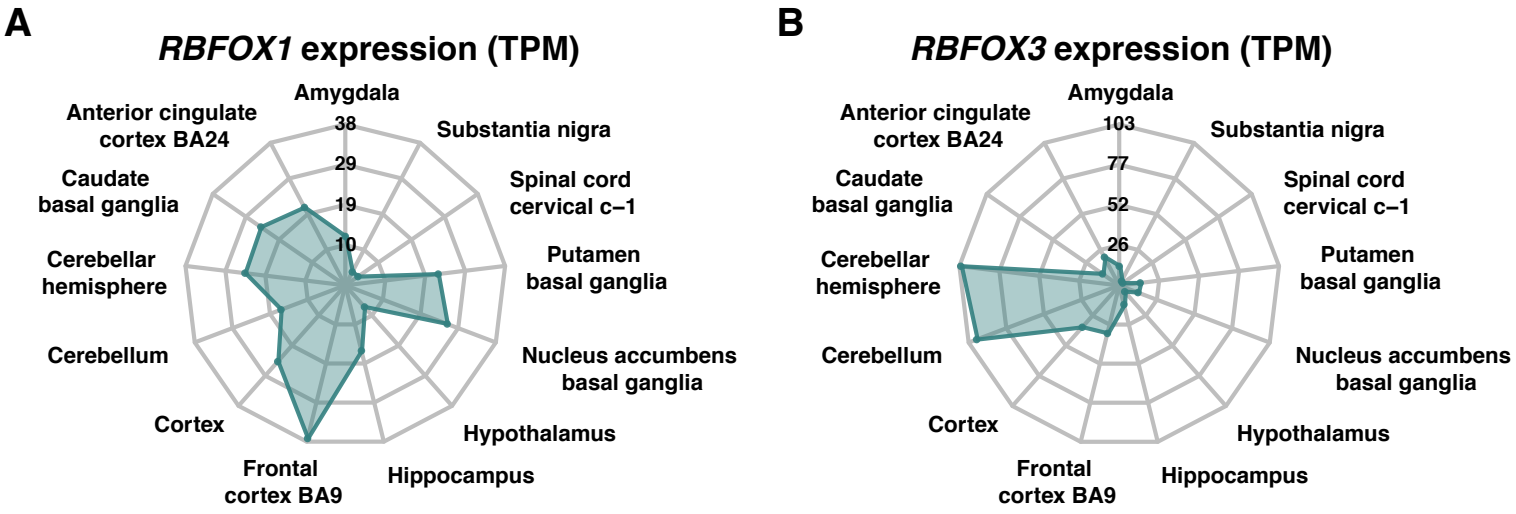


Figure S13. Gene Expression Levels of *RBFOX1* and *RBFOX3* across 13 Brain Regions
(A), (B) Radar plots showing mean gene expression levels (TPM) of two RBPs (A: *RBFOX1*; B: *RBFOX3*) in each brain region.

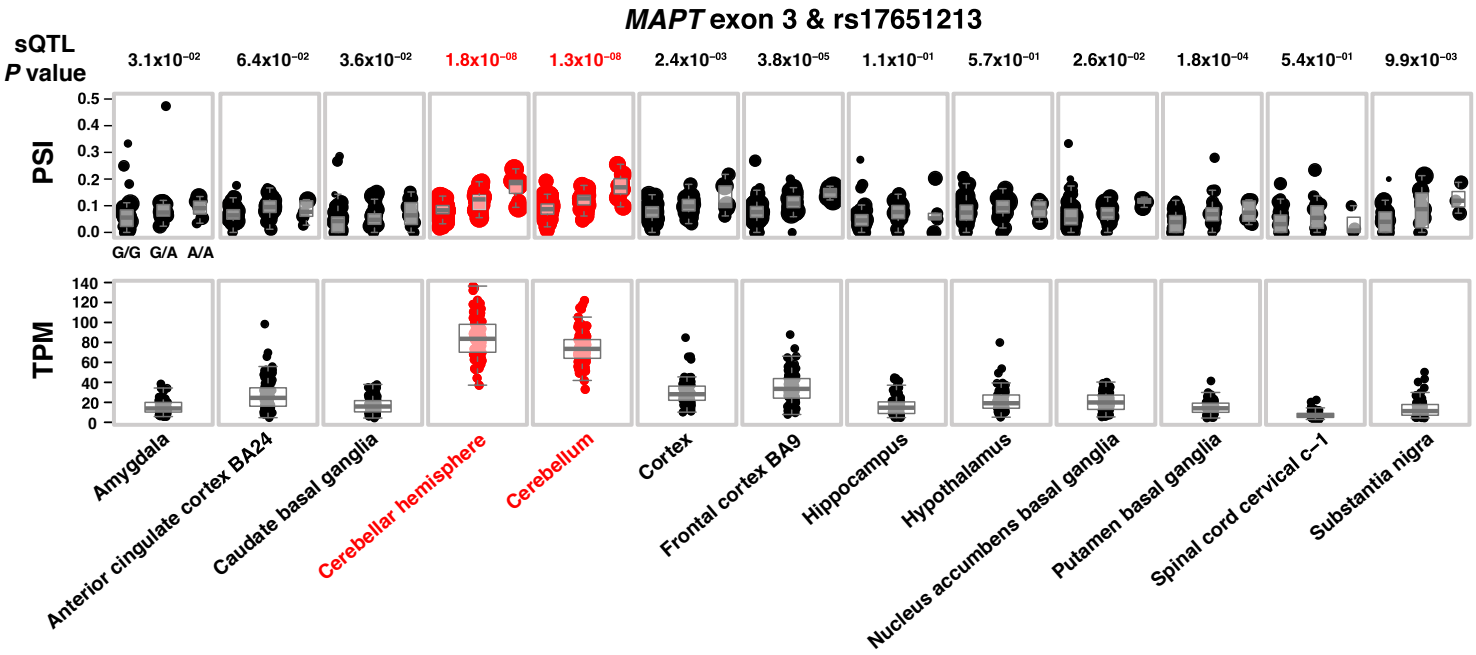


Figure S14. Regionally Specific sQTL (*MAPT* Exon 3 and rs17651213) and *RBFOX2* Gene Expression

(Upper panel) Box plots showing association between *MAPT* exon 3 and rs17651213 in each brain region. Brain regions where the association is significant are shown in red. (Lower panel) Box plots showing gene expression level of *RBFOX2* in each brain region. Each dot represents one tissue sample.

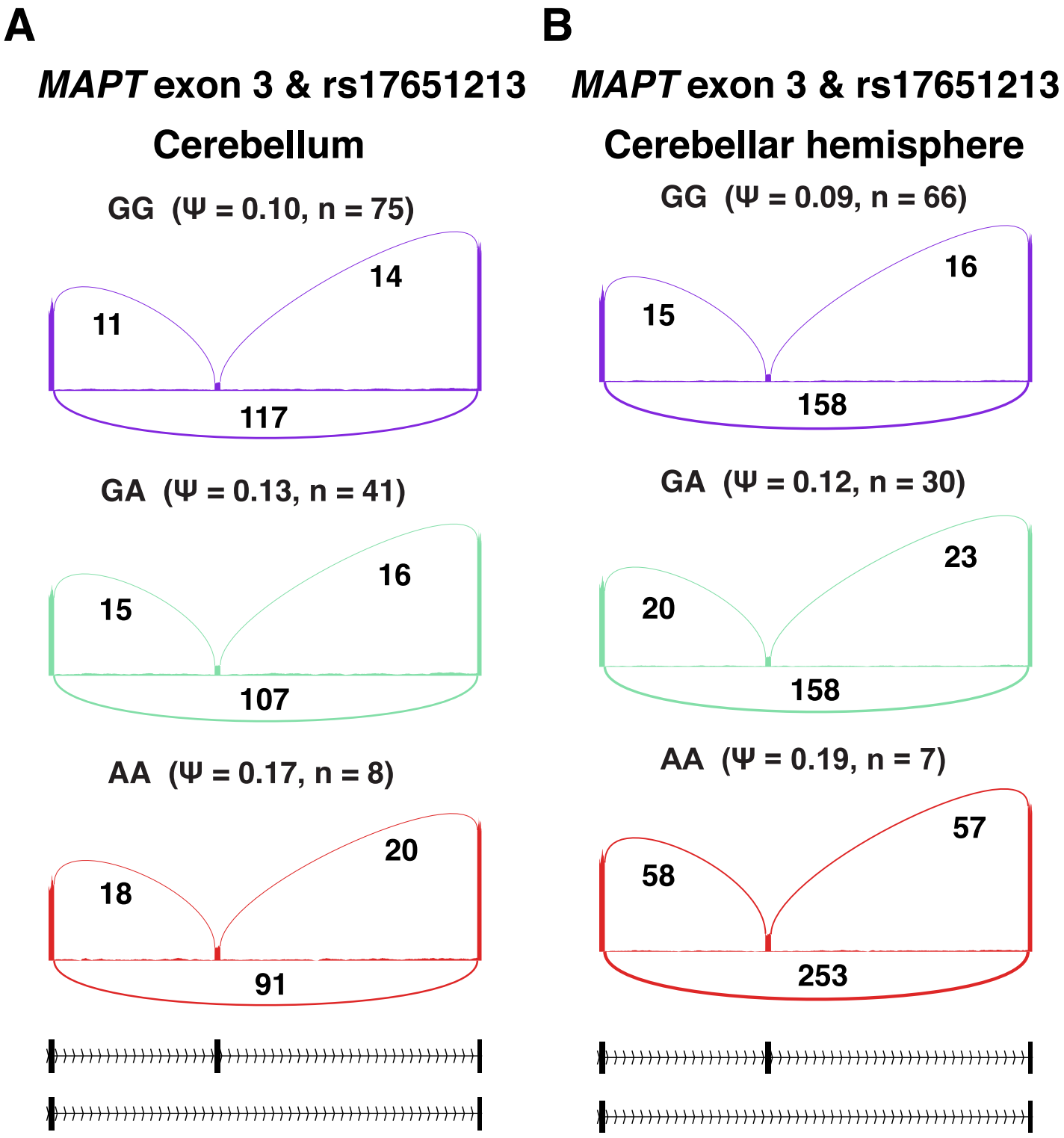


Figure S15. Sashimi Plots of *MAPT* Exon 3 and rs17651213 in Cerebellum and Cerebellar Hemisphere

(A), (B) Sashimi plots of the sQTL (*MAPT* exon 3 and rs17651213) in two significant brain regions (A: cerebellum; B: cerebellar hemisphere).

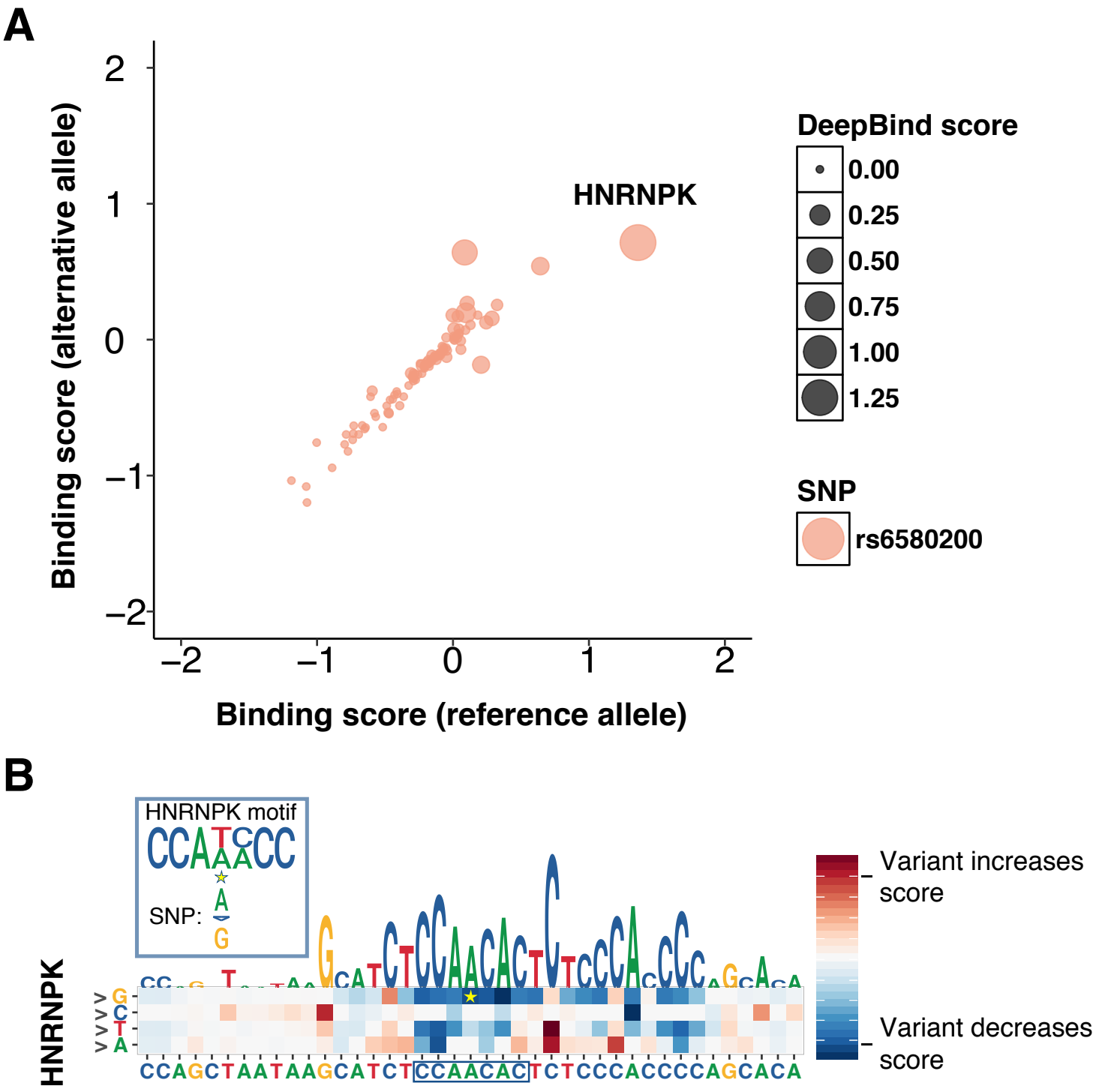


Figure S16. SNP rs6580200 May Regulate Splicing of *CXXC5* Exon 2 by Affecting HNRNPK Binding
(A) Bubble plot showing effects of one SNP (rs6580200) on RBP-RNA binding, as predicted by DeepBind. Axes show RBP binding scores of sequences with reference allele (X-axis) or alternative allele (Y-axis) for each RBP. Bubble size is proportional to the difference in DeepBind scores between the two alleles. (B) DeepBind variant map showing SNP rs6580200 with the HNRNPK binding site. Star indicates the position of the SNP.

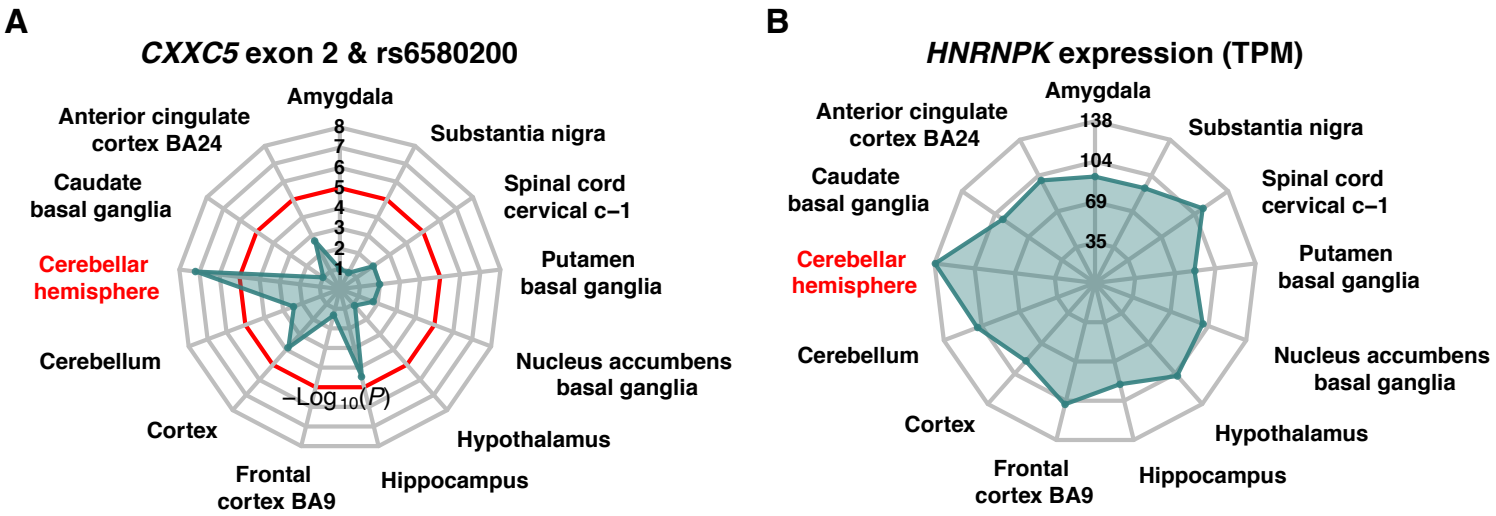


Figure S17. Significance of the sQTL (*CXXC5* Exon 2 and rs6580200) and the Gene Expression Level of *HNRNPK* across 13 Brain Regions

(A) Radar plot showing significance ($-\log_{10}(P$ value)) of the sQTL (*CXXC5* exon 2 and rs6580200) in each brain region. (B) Radar plot showing mean gene expression level (TPM) of the RBP (*HNRNPK*) in each brain region.

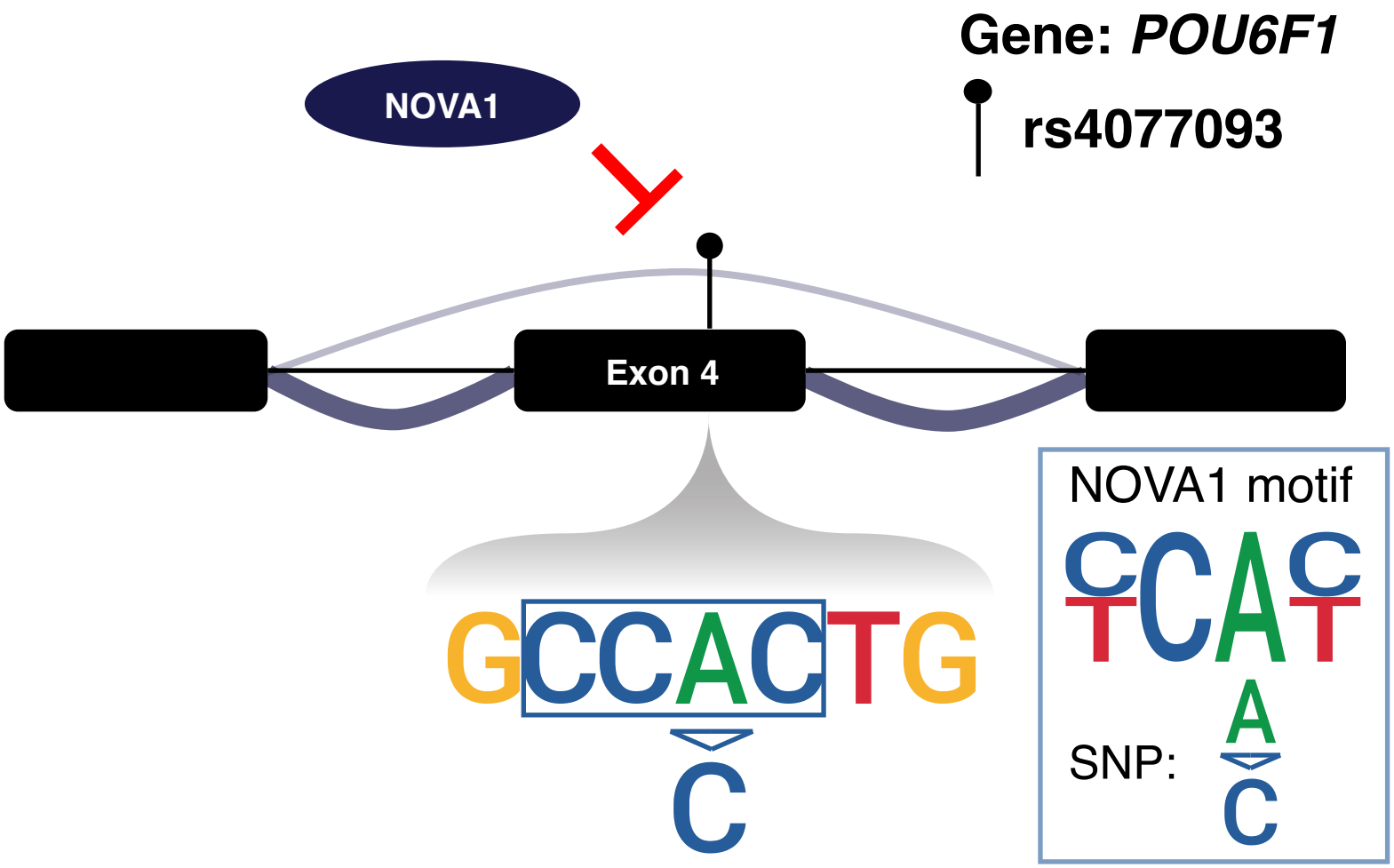


Figure S18. SNP rs4077093 May Regulate Splicing of *POU6F1* Exon 4 by Affecting NOVA1 Binding
SNP rs4077093 (A>C) potentially increases inclusion of *POU6F1* exon 4 by disrupting a NOVA1 binding site within the exon.

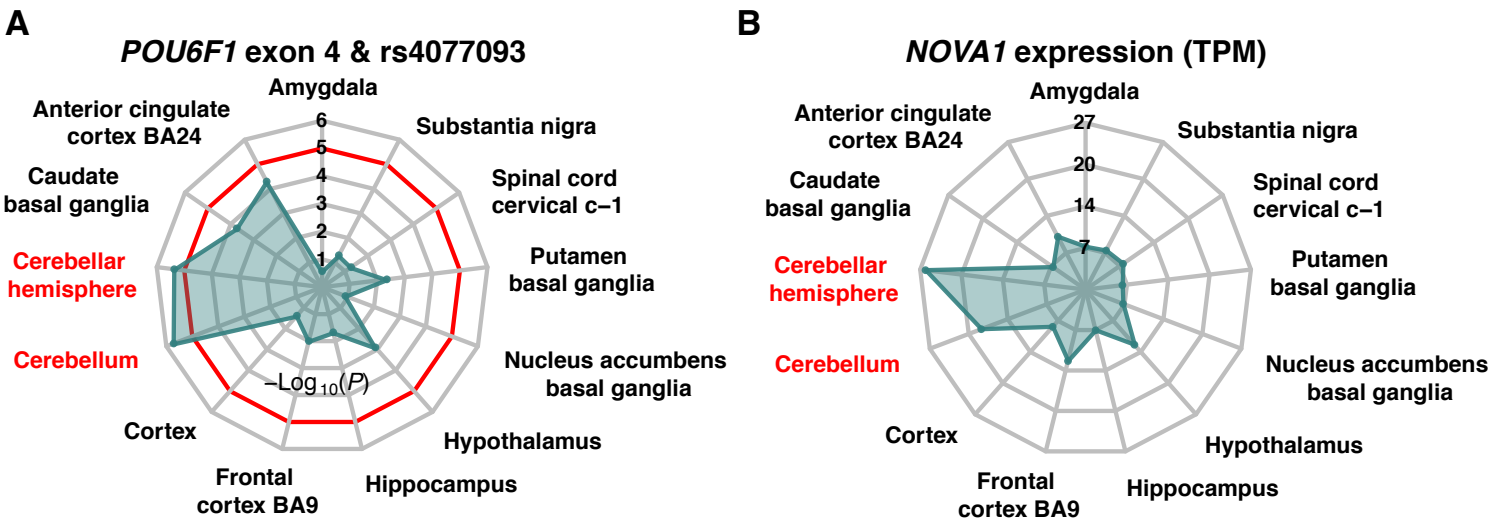


Figure S19. Significance of the sQTL (*POU6F1* Exon 4 and rs4077093) and the Gene Expression Level of *NOVA1* across 13 Brain Regions

(A) Radar plot showing significance (-$\log_{10}$(*P* value)) of the sQTL (*POU6F1* exon 4 and rs4077093) in each brain region. (B) Radar plot showing mean gene expression level (TPM) of the RBP (*NOVA1*) in each brain region.