Genetic influences on cell type specific gene expression and splicing during neurogenesis elucidate regulatory mechanisms of brain traits

Nil Aygün[1,2], Angela L. Elwell[1,2], Dan Liang[1,2], Michael J. Lafferty[1,2], Kerry E. Cheek[1,2], Kenan P. Courtney[1,2], Jessica Mory[1,2], Ellie Hadden-Ford[1,2], Oleh Krupa[1,2], Luis de la Torre-Ubieta[3-6], Daniel H. Geschwind[3-6], Michael I. Love[1,7], Jason L. Stein[1,2,8]

1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

2) UNC Neuroscience Center University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

3) Neurogenetics Program, Department of Neurology, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA 90095, USA

4) Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA 90095, USA

5) Department of Human Genetics, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA 90095, USA

6) Department of Psychiatry and Biobehavioral Sciences, Semel Institute, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA 90095, USA

7) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

8) Lead Contact: jason_stein@med.unc.edu

Summary

Interpretation of the function of non-coding risk loci for neuropsychiatric disorders and brain-relevant traits via gene expression and alternative splicing is mainly performed in bulk post-mortem adult tissue. However, genetic risk loci are enriched in regulatory elements of cells present during neocortical differentiation, and regulatory effects of risk variants may be masked by heterogeneity in bulk tissue. Here, we map e/sQTLs and allele specific expression in primary human neural progenitors (n=85) and their sorted neuronal progeny (n=74). Using

colocalization and TWAS, we uncover cell-type specific regulatory mechanisms underlying risk for these traits.

Introduction

Genome wide association studies (GWAS) have identified many common non-coding variants associated with risk for brain structure, neurodevelopmental disorders, and other brain-related traits[1–7]. However, it is challenging to determine the mechanism of non-coding variants because, in general, (1) the genes impacted by non-coding risk variants are unknown, (2) the cell type(s) and developmental period(s) where the variants have an effect are not known, (3) and there may be limited availability of tissue representing the causal developmental stage and cell type.

One potential mechanism by which non-coding genetic variation can influence brain traits is through alterations in gene expression, or expression quantitative trait loci (eQTLs). Genetic variation also impacts transcript splicing[8–10], and several studies have implicated genetically mediated alterations in splicing as important risk factors for neuropsychiatric disorders[11–13].

Most current efforts to explain the function of these risk loci rely on mapping local expression and splicing quantitative trait loci (e/sQTLs) in bulk adult brain tissue[14,15]. However, genetic risk loci are enriched in cell types relevant for neocortical differentiation that are not present in the adult brain[16,17]. e/sQTL studies have been performed on human fetal brain bulk cortical tissue[18–20], but there are advantages to a cell type specific approach given that regulatory effects of risk variants may be masked by heterogeneity in bulk tissue[21,22].

Utilizing a cell type specific *in vitro* model system including neural progenitors (N_{donor} = 85) and their virally labelled and sorted neuronal progeny (N_{donor} = 74) derived from a multi-ancestry population, here we investigated how common genetic variants impact brain related traits through gene expression and splicing during human neurogenesis. We discovered 2,079/872 eQTLs in progenitors and neurons, and 5,676/4,590 sQTLs in progenitors and neurons,

respectively. Importantly, 66.1%/47% of eQTLs in progenitor/neuron and 78.8%/76.1% sQTLs in progenitor/neuron were unique, and were not found in fetal bulk brain e/sQTLs from a largely overlapping sample[19] and in adult bulk e/sQTL data from GTEx project[23]. We showed both eQTLs and sQTLs colocalized with known GWAS loci for neuropsychiatric disorders and other brain relevant traits in a cell type specific manner. By integrating the dataset generated here with cell type specific chromatin accessibility from the same cell lines[24] and brain structure GWAS[5], we propose a regulatory mechanism whereby genetic variation influences a proxy of human intelligence across multiple levels of biology. Furthermore, we genetically imputed disease/brain trait susceptibility gene expression and alternative splicing in these cell types using transcriptome-wide association study (TWAS), that identified cell type and temporal specific risk genes and introns.

Results

Transcriptomics of primary human progenitors and neurons

We established an *in vitro* culture of primary human neural progenitor cell (phNPC) lines derived from genotyped human fetal brain tissue (N = 89 unique donors) at 12-19 post conceptional weeks (PCW) (14-21 gestation weeks), that recapitulates the developing human neocortex[25–28] (Figure 1A, Methods). Immunofluorescence of the cells showed that undifferentiated progenitors expressed PAX6 and SOX2 (90-95%), consistent with a homogenous culture of radial glia[29,30] (Figure 1A). At 5 weeks post-differentiation, phNPC cultures were transduced with a virus which expresses EGFP in neurons (AAV2-hSyn1-EGFP), which enabled us to isolate neurons via FACS sorting at 8 weeks post-differentiation (Figure 1A, Supplementary Figure 1A-B, Methods).

We acquired transcriptomic profiles of progenitors and neurons via RNA-sequencing (Methods; Supplemental Figure 1C), where we observed that gene expression profiles were more

correlated between libraries generated from the same donor thawed at different times as compared to libraries across different donors (Supplementary Figure 2A). After correction for technical confounds (Supplementary Figure 2B), progenitors and neurons clustered separately by principal component analysis (PCA) of global gene expression, indicating global transcriptomic differences by cell type (Figure 1B).

Assessing fidelity of phNPC cultures

Next, we evaluated cell type specificity of progenitors and neurons by evaluating the expression of literature based marker genes[31]. Progenitors exhibited higher expression of ventricular radial glial (*TAGLN2*), outer radial glial (*HOPX*), and pan radial glial markers (*SOX2*) (Figure 1C). Conversely, neurons had higher expression of known neuronal markers (*CTIP2* and *RBFOX3*).

In order to identify gene expression differences between each cell type, we performed a differential gene expression (DGE) analysis. We found 1,281 genes upregulated in progenitors, and 2,276 genes upregulated in neurons ($|\log2FC| > 1.5$ and $FDR < 0.05$; Figure 1D, Supplementary Table 1). We identified genes involved in cell cycle and neurotransmission upregulated in progenitors and neurons, respectively (Figure 1D). Gene ontology confirmed enrichment of differentially expressed genes in these expected pathways (Figure 1E).

We next evaluated how well the *in vitro* progenitors and neurons we generated model *in vivo* neurodevelopment. We implemented the transition mapping (TMAP) approach for a global assessment of transcriptomic overlap between *in vitro* cultures and *in vivo* post-mortem human brain samples, as described in our previous work[26] (Methods). We compared the transition from progenitor to neurons with laser capture microdissection of cortical laminae from postmortem human fetal brain at 15-21 PCW[32]. We observed the strongest overlap in the transition from progenitors to neurons with the transition from outer subventricular zone (oSVZ) to intermediate

zone (IZ) or subplate zone (SP) (Figure 1F). Likewise, we also performed TMAP analysis for different fetal developmental periods[15,33]. We also observed the highest match with the transition from 8-9 PCW to 13 PCW or 16-17 PCW (Supplementary Figure 3). These results support the *in vivo* fidelity of our culture system representing neurogenesis during mid-fetal development.

Cell type specific local expression quantitative loci (eQTL) analysis

To investigate the impact of genetic variation on gene expression, we performed a local eQTL analysis by testing the association of each gene's expression levels with genetic variants residing within +/-1 Mb window of its transcription start site (TSS)[34,35] (Figure 2A, see Methods) . We implemented a linear mixed effects model to stringently control for population stratification using a kinship matrix as a random effect with inferred technical confounders as fixed effects, separately for each cell type (λ_{GC} for progenitor = 1.028 and λ_{GC} for neuron = 1.007; see Methods, Supplementary Figure 2G). After retaining associations which were lower than 5% false discovery rate with a hierarchical multiple testing correction[36,37] (Methods), we obtained conditionally independent eQTLs (Supplementary Figure 4A and 4B, see Methods). We identified 1,741 eGenes with 2,079 eSNP-eGene pairs in progenitors and 840 eGenes with 872 eGene-eSNP pairs in neurons (Supplementary Table 2). 54.2%/75.8% of progenitor/neuron primary eQTLs were shared with m-value > 0.9 (Supplementary Figure 5A-B). In agreement with previous local eQTL studies[19,35,38–40], the distribution of primary eSNPs relative to TSS revealed that index eSNPs with higher significance were often closer to the TSS of the eGene that they regulate for both cell types (Figure 2B). We found that 73.6%/75.9% of primary eSNPs were within 100 kb of the TSS, and 13%/12.7% of primary eSNPs were in the promoter of the eGene (2kb upstream of TSS) in progenitors/neurons, respectively. eGenes were mostly (70.8%/54.5%) protein-coding, though lincRNAs were detected as well (6.5%/11.6%) in progenitors/neurons, respectively (Figure 2C).

We determined if eSNPs were enriched (see Methods) in chromatin states of the fetal human brain[17,41] (Figure 2D, Methods). Both progenitor and neuron specific eSNPs were enriched in promoters, enhancers and actively transcribed sites present in the fetal brain, but depleted mostly in quiescent chromatin regions. We also measured enrichment of cell type specific chromatin accessibility QTLs within cell type specific eQTLs[16] and bulk fetal eQTLs[19] from a partially overlapping sample, as well as adult bulk cortex eQTLs[23] (Figure 2E). We observed that variants associated with chromatin accessibility within a peak (caSNP) in progenitors were enriched most strongly in progenitor eQTLs. Neuron caSNPs showed a similar pattern, suggesting that genetic variants may alter the regulatory landscape in a parent cell type, leading to expression changes only after differentiation. These observations supported the notion that genetic variation within functional regulatory elements in the tissue and cell-type of interest impact gene expression.

Exploring cell type and temporal specificity of eQTLs

We aimed to determine the utility of our cell type specific eQTL study by comparison to pre-existing bulk brain eQTL studies. Comparing our results to a bulk fetal cortical wall eQTL dataset from a previous study using a partially overlapping set of donors[19], we observed 26%/45% of eSNP-eGene pairs in progenitors/neurons were common with bulk fetal cortical eQTL data, where we considered genetic variants with pairwise LD $r^2 > 0.8$ as marking same loci between two datasets (Figure 2F, odds ratio test p-value: 5.06e-23). This indicates that both cell type specific eQTLs, but especially progenitor eQTLs in our cell type specific system offer novel regulatory mechanisms which can provide additional information beyond existing prenatal datasets[18–20].

We next explored the temporal specificity of cell type specific eQTLs by utilizing adult brain bulk cortical eQTL data from the GTEx project[23]. We observed 18.9%/28.3% of eSNP-eGene pairs in

progenitors and neurons, respectively, were also found in adult brain eQTL data (Supplementary Figure 5C). That suggests largely independent genetic mechanisms regulating genes from development to adulthood[20].

Cell type specific allele specific expression

We performed cell type specific allele specific expression (ASE) analysis to understand allelic effects on gene expression within heterozygous individuals as a complementary approach to eQTL mapping[35,42] (Figure 3A). We found 6,402 and 8,698 ASE sites in progenitors and neurons, respectively (Figure 3B, see Supplementary Figure 6A-B for MA plot, Supplementary Table 3). Next, we determined the number of eQTLs supported by ASE sites. We observed that 53% of progenitor and 45.5% of neuron eGenes (with comparable filtering criteria with ASE) overlapped with genes that had ASE sites (see Methods, Figure 3C). Importantly, 32.7%/34.3% of progenitor/neuron specific eQTLs, respectively, were supported by ASE, indicating that we could detect genetic variants altering gene expression with an independent within-donor method that is not subject to cross donor technical confounding, like population stratification[43]; Figure 3D). Both progenitor and neuron eQTLs shared with genome-wide significant ASE sites were highly concordant in effect size and direction (progenitor, r=0.83; neuron, r=0.849), while the eQTLs shared with nonsignificant ASE sites were still concordant but to a lesser degree (progenitor, r=0.54; neuron, r=0.517) (Figure 3E).

We also identified allele specific effects supporting eQTLs colocalized with GWAS for neuropsychiatric disorders. For example, we found a progenitor eSNP (rs1051168) regulating *NMB* gene expression (p-value: 2.78e-09) colocalized with a SCZ index SNP (rs12908161) (see Methods for colocalization approach, Figure 3F). Variant rs1051168 was also supported by an ASE site (FDR = 1.01e-8), where the risk allele for SCZ (G) was associated with decreased *NMB* expression. NMB gene is a bombesin-related peptide highly expressed in the central

nervous system[44] and was associated with schizophrenia risk through a previous fetal brain eQTL based genetic imputation[45,46]. Our observation proposes a cell type specific eQTL mechanism that was also supported by allele specific *NMB* expression to understand the regulatory mechanism underlying a schizophrenia GWAS risk locus.

Cell-type specific splicing quantitative trait loci (sQTL)

Given the impact of genetic variation on alternative splicing[9,11,19,47], we next performed a splicing quantitative loci (sQTL) analysis separately within progenitors and neurons. We quantified alternative intron excisions as percent spliced in (PSI) by implementing the Leafcutter software, an annotation free approach that allows for discovery of novel isoforms[48]. We found 34,449/35,285 intron excisions present more often in progenitors/neurons, respectively (see Methods, Supplementary Table 4). As a specific example, we found a differential alternative splicing site within the *DLG4* gene encoding the postsynaptic density protein 95 (PSD-95). An exon skipping splice site supporting nonsense mediated decay (splice 1, ENST00000491753) was upregulated in progenitors; while another splice site supporting multiple protein coding transcripts (splice 2) was upregulated in neurons (Figure 4B). Post-transcriptional repression of PSD-95 expression in neural progenitors via nonsense mediated decay at splice 1 site has been previously experimentally validated[49,50], giving strong confidence in the cell type specific splicing calls.

For the sQTL analysis, we implemented an association test between PSI of each intron excision and genetic variants located within a ± 200 kb window from the start and end of the splice junctions (Figure 4A). We retained significant associations which were lower than 5% false discovery rate by implementing a hierarchical multiple testing correction (see Methods), and applied conditional analysis to identify independent sQTLs. We identified 4,708 intron excisions associated with 5,676 sSNPs-intron junction pairs in progenitors and 4,039 intron excisions

associated with 4,590 sSNPs-intron junction in neurons (Supplementary Figure 7A-B, 35.6%/79.7% of progenitor/neuron primary sQTLs were shared with m-value > 0.9, Supplementary Table 5). We observed that genetic variants closer to splice junctions exhibited a greater impact on splicing (44.4% of primary progenitor sSNPs and 42.4% of primary neuron sSNPs were within 10kb of the splice junctions, Figure 4C). As an example of an sSNP within a splice junction, we found that the indel variant rs11382548 creates a canonical splice acceptor sequence and then impacts two different intron excisions supporting alternative 3' splice sites (Figure 4D) for *TMEM216*. Deletion of the A nucleotide at a canonical splice acceptor site of the last exon of *TMEM216* leads to disruption of the alternative splicing event for transcript ENST00000334888, and increased usage of transcript ENST00000398979 and ENST00000515837 in both progenitors and neurons. This splice variant may be relevant to neurogenesis because knockdown reduces division of both apical and intermediate progenitor cells during corticogenesis[51].

Interestingly, many splice sites were previously unannotated in the gene models we used (Ensembl Release 92). We detected 8.8%/11.3% cryptic at the 5' end; 12%/12% cryptic 3'end; 8.9%/10.6% both cryptic ends for intron excision within progenitors/neurons. We next sought to determine if sQTLs are enriched in known functional genomic elements. We observed sQTLs show a strong and significant enrichment in splice sites, introns and 5'/3' UTR sites (Figure 4E).

We also found that cell type specific sQTLs in progenitors and neurons were enriched for RNA binding sites from CLIP-seq databases of 74/76 different RNA-binding proteins (RBP)[52] (Figure 4F, Supplementary Table 6). Strikingly, 22/24 of those proteins were specifically enriched in progenitor and neuron specific sQTLs, respectively. Among RBP binding sites specifically enriched for progenitor sQTLs, we found LIN28B, known to play a role in neural progenitor proliferation and differentiation[53]. On the other hand, for neurons, we detected enrichment of the

NPM1 regulating neuronal survival[54]. These observations suggest that sQTLs alter the function of RBPs with cell type specific splicing roles during neural development.

In order to examine if variants associated with alternative splicing also alter expression of the same genes, we compared cell type specific sQTLs with cell type specific eQTLs. Only 17.6%/9.5% of sGenes, the genes that harbor intron excisions, were also eGenes for progenitors and neurons eQTLs, respectively (Supplementary Figure 7C, upper panel). Furthermore, we also found that only 2.9% and 1.4% of sGene-sSNP pairs overlapped with eGene-eSNP pairs for progenitors and neurons, respectively (Supplementary Figure 7C, lower panel), indicating that sQTLs generally function through independent mechanism from eQTLs.

We next examined the impact of cell-type specificity on identification of sQTLs (see Methods; Figure 4G). We found 33%/38% of intron excisions, and 17%/23.5% of sSNP-intron excision pairs in progenitor and neuron specific sQTLs, respectively were shared with bulk cortical fetal tissue sQTLs (Figure 4G). We observed a smaller overlap of progenitor sQTLs with bulk cortical fetal tissue as compared to neuron sQTLs indicating that cell-type specificity allowed for novel discovery of progenitor sQTLs (Figure 4G, odds ratio test p-value: 1.7e-15). Also, we found 6%/2.4% of sSNP-intron junction pairs in progenitors and neurons, respectively shared with adult brain bulk cortical sQTL data from GTEx project[23] (Supplementary Figure 7D), showing temporal specificity of cell type specific sQTLs.

Regulatory mechanisms of brain related GWAS

We sought to explain the regulatory mechanism of individual loci associated with neuropsychiatric disorders, brain structure traits, and other brain-relevant traits by leveraging genetic variants regulating cell-type specific gene expression and splicing. We co-localized GWAS loci of these traits with cell-type specific eQTLs and sQTLs using a conditional analysis

to ensure the loci were shared across traits[55] (Supplementary Figure 8A; see Methods for the list of GWAS used for this analysis).

We discovered 41, 13, and 20 GWAS loci that co-localized specifically with progenitor eQTL, specifically with neuron eQTLs, or with both cell types, respectively (Figure 5A, Supplementary Table 7). These observations show that the same genetic variants impact gene expression, neuropsychiatric traits, and brain structure in a cell type specific manner. Importantly, 98 colocalized loci-gene pairs per trait were not found using fetal bulk cortical tissue eQTLs where tissue heterogeneity may have masked their detection (Figure 5B).

Next, we leveraged our cell-type specific chromatin accessibility QTL (caQTL) dataset[24] together with eQTLs in order to explain the regulatory mechanism underlying GWAS loci associated with brain relevant traits. As a specific example, we found a colocalization of a locus within the *CENPW* gene across caQTLs, eQTLs, GWAS for Global Surface Area (GSA) and for Educational Attainment (EA) (Figure 5C). The progenitor index eSNP rs4897179 that was not detected in bulk cortical fetal tissue eQTLs (nominal p-value = 3.26e-07 in progenitors, nominal p-value = 0.068 in neurons, and nominal p-value = 0.08 in fetal cortical bulk tissue), for the *CENPW* eGene, was colocalized with variant rs9388490, which is the index SNP for both GSA and EA GWAS (nominal p-value = 4.954e-12 in GSA GWAS, and nominal p-value = 1.43e-08 in EA GWAS). Also, we found that a SNP (rs9388486) located within a chromatin accessible peak region 107 bp upstream of TSS of the *CENPW* gene was colocalized with the index eSNP. We therefore consider rs9388486 as the potential causal variant, and noted that the C allele disrupts the motifs of the transcription factors CREM, ATF2, ATF4 and ATF1 (Figure 5D). CENPW is required for appropriate kinetochore formation and centriole splitting during mitosis[56], and increased CENPW levels lead to apoptosis in the developing zebrafish central nervous system[57]. Overall, these observations propose a cell type specific mechanism whereby the C

allele at variant rs9388486 disrupts transcription factor binding and diminishes accessibility at the *CENPW* gene promoter, resulting in decreased *CENPW* gene expression levels in progenitors, presumably altering neurogenesis or reducing apoptosis, leading to increased cortical surface area and higher cognitive function.

We also aimed to examine cell type specific splicing QTLs colocalized with GWAS loci. We observed 28, 23, and 29 GWAS loci in total that co-localized with specifically progenitor/neuron sQTLs and sQTLs present in both cell types (Figure 6A, Supplementary Table 7). Similar to eQTL colocalizations, we observed that 124 different colocalized loci-intron junction-trait combinations were detected only with cell type specific sQTL, but not fetal cortical bulk sQTLs (Figure 6B). Interestingly, we detected a progenitor sSNP (rs3740400, nominal p-value: 3.29e-10) associated with an unannotated exon skipping splicing event for the *AS3MT*, that was not detected in fetal bulk sQTL data (nominal p-value: 0.005), was colocalized with an index GWAS SNP for schizophrenia[1] (rs11191419) (Figure 6C). The risk allele for schizophrenia (T) was associated with more usage of this splicing site. A transcript supported by the same exon skipping event was discovered previously in the adult brain, that was regulated by variant rs7085104 in LD with rs3740400[58]. Here, we demonstrate genetically regulated upregulation of this transcript in neural progenitors, but not early born neurons, which shows a novel developmental basis for this previously identified association. We also observed that another progenitor specific sSNP (rs1222218) regulating a novel alternative exon skipping event for *ARL14EP* gene, was colocalized with SCZ index SNP (rs1765142) (Figure 6D). The risk allele for SCZ led to more frequent skipping of the exon, supporting expression of a novel isoform. *ARL14EP* gene has been shown to play a role in axonal development in the mouse neurons[59]. Here, we propose a novel transcript of this gene with expression in progenitors as a risk factor for SCZ.

Genetic imputation of cell type specific GWAS susceptibility genes and alternative splicing

Next, we imputed genes and alternative splicing associated with brain related traits by integrating the polygenic impact of cell type specific regulatory variants with GWAS risk variants in a transcriptome-wide association study (TWAS) approach[60,61]. We found 1,703/973 genes and 6,728/6,799 intron junctions as significantly cis-heritable in progenitors/neurons (heritability p-value < 0.01). Given the difference in population structure between our dataset and European neuropsychiatric GWAS, we tested the impact of different LD estimates on the TWAS results (see Methods; Fig 7A and 7C Supplementary Figure 9A,B). We observed a high overlap between these two analyses with the majority of genes/introns detected being shared regardless of LD structure (Supplementary Figure 9C). We found the cis heritable impact of 124/102 genes in progenitor/neuron and 359/365 intron junctions in progenitor/neuron significantly correlated with at least one neuropsychiatric disorders or other brain related-traits (Supplementary Table 8). Of those significant TWAS genes/introns, to separate conditionally independent genetic predictors from the co-expressed ones, we performed a conditional analysis[62,63], and defined those predictors as jointly independent. We performed cell-type specific TWAS on both gene expression and splicing for schizophrenia (jointly independent genes: 23 progenitor, 26 neuron; jointly independent introns: 59 progenitor, 64 neuron), IQ (jointly independent genes: 25 progenitor, 24 neuron; jointly independent introns: 38 progenitor, 65 neuron), and neuroticism (jointly independent genes: 13 progenitor, 15 neuron; jointly independent introns: 40 progenitor, 32 neuron) (Figure 7A and Supplementary Figure 10A; see Supplementary Figure 9D for the full list of traits tested). Also, for each trait, we found novel loci not discovered in colocalization analysis, indicating that the polygenic effect of multiple loci on gene expression/splicing provides additional power to identify genetically regulated genes compared to a single-marker testing approach.

We next explored the cell type and temporal specificity of our TWAS approach for genes and introns junctions by performing TWAS analysis by using weights calculated from bulk cortical fetal tissue[19], and adult brain eQTLs from the Common Mind Consortium[14,63] (CMC) (Supplementary Figure 9D and Supplementary Table 9 for performance comparison). Most TWAS findings were specific to a cell type or temporal e/sQTL dataset, rather than broadly detected, indicating that different developmental or cell type e/sQTL datasets contribute complementary information about genes influencing risk for neuropsychiatric disorders or other brain traits. We observed an example supporting the cell type and temporal specificity of our analysis at the *B3GALNT2* gene locus using TWAS for IQ (Figure 7B). Despite IQ GWAS falling short of the genome-wide significance threshold at this locus, we detected that genetically imputed expression of the *B3GALNT2* gene in progenitors was significantly correlated with IQ. Importantly, neither IQ TWAS in fetal bulk tissue nor in CMC adult brain tissue were found to be significantly associated with *B3GALNT2* gene expression (Figure 7B). The gene was also not significantly cis-heritable within neurons. Mutations in the *B3GALNT2* gene play a role in glycosylation of α-dystroglycan and were associated with intellectual disability in individuals with congenital muscular dystrophy[64,65]. Overall, here we showed that an increase in genetically imputed *B3GALNT2* gene expression in progenitors is associated with lower IQ, suggesting this gene's early cell type specific impact on cognitive function.

Within the cell-type specific splicing TWAS, for *MRM2* gene, we found an intron junction more frequently spliced that was associated with increased risk for schizophrenia specifically in progenitor cells (TWAS-Z: 6.538), but it was not significantly cis-heritable within neuron, fetal bulk or adult bulk data (Supplementary Figure 10B). MRM2 is a mitochondrial rRNA methyltransferase[66], and was found to be associated with intellectual disability[67] and mitochondrial encephalopathy[66]. Here, we propose a cell type specific developmental basis for alternative splicing of the *MRM2* gene associated with risk for schizophrenia.

Discussion

Here, we investigated the influence of genetic variation on brain related traits within a cell type specific model system recapitulating a critical time period of human brain development, neurogenesis. Our analysis discovered features of gene regulation that will be complementary to previous eQTLs and sQTLs identified in bulk human brain in that: (1) we identified thousands of novel eQTLs, ASEs, and sQTLs during brain development that are enriched in regulatory elements present during neurogenesis; (2) most e/sQTLs in progenitors/neurons were not identified in previous bulk tissue post-mortem tissue datasets from fetal bulk cortical tissue indicating the importance of cell-type specificity for identifying genetic influences on gene regulation; (3) using this resource, we are able to propose cell-type specific variant-gene/transcript-trait(s) pathways to further explore molecular and developmental causes of neuropsychiatric disorders; (4) by integrating the polygenic effects across traits and gene expression, we are able to impute cell type specific gene expression/alternative splicing dysregulation in individuals with neuropsychiatric disorders in time periods prior to disease onset.

As one example of a cell-type specific variant-gene-trait pathway, we discovered a locus near the *CENPW* gene colocalized across cell-type specific caQTL, eQTL, brain size, and cognitive function. Through the integration of multi-omic gene-to-trait databases, we hypothesize that the C allele at rs9388486 leads to decreased TF binding of up to 4 transcription factors (ATF1/2/4, CREM) in progenitors, resulting in decreased chromatin accessibility at the promoter peak, decreased expression of *CENPW*, leading to increased cortical surface area, and increased cognitive function. The *CENPW* gene has a strong role in proliferation, as it is required for kinetochore formation during mitosis[68]. This is consistent

with progenitor proliferation influencing surface area, as described in the radial unit hypothesis[69]. Increased levels of *CENPW* may cause death of progenitor cells either by directly being an apoptotic inducer or by triggering apoptosis in response to an imbalance in cell homeostasis with excessive mitotic activity[57]. In all, we demonstrate how integration across multi-level biological data can be used to propose functional mechanisms underlying complex traits, and future studies may be able to develop computational models to propose causal pathways across multi-omic QTL data[9,70,71]. Such information will be crucial to both design efficient functional validation experiments as well as to leverage GWAS loci to advance treatment targets for neuropsychiatric disorders.

Though the most commonly proposed regulatory mechanism by which non-coding genetic variation influences complex traits is through gene expression levels[34], our data also support mechanisms by which genetic variants associated with cell-type specific alternative splicing influence complex brain-relevant traits. Importantly, we observed sQTLs impacting previously unannotated cell-type specific alternative splicing events that are also colocalized with brain relevant GWAS. For example, we found a progenitor specific sSNP regulating one unannotated exon skipping splice site for the *AS3MT* gene. Previous work has shown that a transcript supported by that splice site was associated with risk for schizophrenia in adult neurons[58]. Here, we propose that this alternative splicing event for the *AS3MT* gene also occurred earlier in development and contributes to schizophrenia risk.

Our cell type specific TWAS analysis identified that alteration in expression of multiple genes and transcripts are associated with risk for different neuropsychiatric conditions. We followed a unique TWAS approach allowing us to explore cell type and temporal specificity

by leveraging existing fetal brain bulk and adult e/sQTLs together with the cell-type specific data we generated here. This type of analysis allows the imputation of the genetically regulated component of differential expression within cell types years prior to disease onset. As such, it allows the knowledge of gene expression differences that cannot be gained from post-mortem tissue of cases versus controls, which must be acquired after diagnosis. This window into developmental gene expression differences may be particularly important to understand disease risk, as these results are not subject to confounding by medication use or the altered experiences of the environment of individuals living with a neuropsychiatric illness[72]. Nevertheless, further support for such data could be gained from iPSC lines modeling early developmental time periods from large populations of cases vs controls.

With our cell-type specific model, we propose how and when genetics influence brain related traits through gene expression and splicing. Future cell type specific eQTLs acquired using flow cytometry or scRNA-seq from the developing post-mortem brain will be useful to validate the *in vivo* impact of genetic variants discovered here using an *in vitro* system. Nevertheless, this *in vitro* system has particular utility in that, in the future it may be used to determine the impact of genetic variation in response to activation of specific pathways or response to environmental stimuli. By pursuing cell-type, temporal, and environmental specificity of eQTLs, we expect that a greater degree of mechanisms underlying risk for neuropsychiatric disorders and brain-relevant traits can be uncovered.

Acknowledgements

This work was supported by NIH (R00MH102357, U54EB020403, R01MH118349, R01MH120125), Brain Research Foundation, and NC TraCS Pilot funding to JLS. DHG was

supported by NIH (R37 MH060233, R01 MH094714, UO1MH116489 and R01 MH110927). The following core facilities were utilized for this project: UNC Neuroscience Center Microscopy Core (P30NS045892), UNC Mammalian Genotyping Core, CGIBD Advanced Analytics Core (NIH grant P30 DK034987), UNC Flow Cytometry Core Facility, UNC Vector Core, UNC Research Computing. Additional core facilities utilized for this project were: UCLA CFAR (5P30 AI028697), and the UCLA Neuroscience Genomics Core. We thank Dr. Karen L. Mohlke and Dr. Yun Li for helpful comments, Dr. Eric Wexler for the idea of phNPC eQTL, and Dr. Stephen Montgomery for clarifying the eigenMT method.

Author Contributions

JLS, DHG, and LTU conceived the study. JLS directed and supervised the study. JLS along with DHG provided funding. ALE, KEC, KPC, MY, LTU, and JLS cultured human neural progenitor cells. ALE performed library preparation. MJL and NA pre-processed the RNA-seq data. DL performed pre-processing of genotyping data and caQTL analysis. JM, EFH and OK performed immunocytochemistry. MIL aided in differential gene expression and allele specific expression analysis methodology. NA managed the integration of the datasets, interpreted results and performed TMAP, differential gene expression and splicing, eQTL, ASE, sQTL, co-localization, enrichment, transcription factor binding motif, and TWAS analyses. JLS and NA wrote the manuscript. All authors commented on and approved the final version of the manuscript.

Declaration of Interests

The authors declare no competing interests.

Data and code availability

Data will be available within to dbGaP upon publication, and codes are available here https://bitbucket.org/steinlabunc/expression_splicing_qtls_public/src/master/.

Figure Legends

Figure 1: Fidelity of primary human neural progenitors (phNPC) and their differentiated neuronal progeny.

(A) Experimental design of the study. Immunolabeling of progenitors and neurons showed that progenitors expressed radial glia markers including SOX2 (90-95%) in red and PAX6 (90-95%) in green, and neurons were positive for EGFP (green) expressed under the control of neuronal specific SYN promoter (scale bar is 100 μm). RNA-sequencing performed for progenitors cells (purple) and neuronal progeny (green).

(B) Principal component analysis of progenitor and neuron transcriptomes indicates cell type specific clustering. Each dot represents a donor from either progenitors (purple) or neurons (green).

(C) Heatmap showing cell type specific expression of literature based progenitor (PAN-RG: Pan-radial glia, V-RG: ventricular radial glia, O-RG: outer radial glia) and neuronal markers listed on the y-axis. The x-axis indicates progenitor (purple) or neuron (green) cells from each donor. The color of heatmap indicates the relative gene expression normalized for each gene between 0 and 1.

(D) MA plot showing differentially expressed genes in progenitor versus neurons: log2FC > 0 and adjusted p-value < 0.05 indicates genes more expressed in neurons shown in green, log2FC < 0 and adjusted p-value < 0.05 indicates genes more expressed in progenitors shown in purple and genes not significantly differentially expressed between two cell types are shown in grey. Blue lines indicates $|\log_2 FC| > 1.5$.

(E) Gene ontology (GO) analysis showing pathways enriched for genes upregulated in progenitors (left, in purple), and for genes upregulated in neurons (right, in green). The x-axis shows adjusted -log10(p-values) for enrichment and each GO term is listed in the y-axis.

(F) Comparison of the transitions between mitotic and postmitotic regions of *in vivo* cortical laminae in the developing cortex and in vitro progenitor and neurons with rank-rank hypergeometric overlap (RRHO) maps. The extent of overlap between *in vivo* and *in vitro* transcriptome was represented by each heatmap colored based on -log10(p) value from a hypergeometric test. Each map shows the extent of overlapped upregulated genes in the bottom left corner, whereas shared downregulated genes are displayed in the top right corners (ventricular zone - VZ; inner and outer subventricular zone - i/oSVZ, intermediate zone - IZ; subplate - SP; inner and outer cortical plate - i/oCP, marginal zone - MZ)

Figure 2: Cell type specific local expression quantitative loci (eQTL) analysis.

(A) A schematic showing that variants within +/- 1MB cis window from the transcription start site (TSS) of each gene were tested for the association with gene expression.

(B) Distance of primary eSNPs from the TSS in progenitors (left) and in neurons (right).

(C) Biotypes of progenitor eGenes in purple and of neuron eGenes in green.

(D) Enrichment of progenitor eSNPs (left), and neuron eSNPs (right) within chromatin states in fetal brain from chromHMM listed on the y-axis. The x-axis shows the effect size of enrichment with 95% upper and lower confidence interval and the plot is color-coded based on -log10(p-value) value from enrichment analysis. Significant enrichments are shown with an asterisk. Enrichment was tested using eQTLs thresholded at the eigenMT-BH p-value.

(E) Enrichment of cell type specific caSNPs within cell type specific, fetal bulk and adult bulk eQTLs with QTLenrich. Adjusted enrichment fold change per data is shown with 95% upper and lower confidence interval. Each dot is colored based on -log10(p-value) and significant enrichments are shown with an asterisk. Dashed blue line indicates x = 1.

(F) Overlap percentage of cell type specific eSNP-eGene pairs shared with fetal bulk eQTLs. Odds ratio test p-value is shown.

Figure 3: Cell type specific allele specific expression analysis.

(A) A schematic illustrating allele specific expression (ASE) in a heterozygous individual for a variant of interest.

(B) Overlap between progenitor and neuron specific ASE sites.

(C) Overlap between eGenes and genes with ASE (progenitors in purple, neurons in green).

(D) Overlap between cell type specific eSNPs and ASE sites (progenitors in purple, neurons in green).

(E) Comparison of the effects of shared ASE sites and eQTLs in progenitors (left in purple) and neurons (right in green). Nonsignificant ASE sites are shown as darker colors for both cell types, and significant ASE sites are shown as lighter colors. Correlation coefficient (r) values are indicated in colors for each category and the red dashed line indicates x = y.

(F) Genomic tracks are color-coded based on LD r^2 relative to the variant rs1051168 at the left showing regional association of variants with schizophrenia (SCZ) and *NMB* gene in progenitor and neurons. Progenitor eSNP rs1051168 associated with *NMB* gene expression was colocalized with SCZ index variant rs12908161 (4th row), that is also a progenitor specific ASE site. Box plots showing VST normalized *NMB* gene expression (upper plot), and allele specific expression (lower plot) across rs1051168 genotypes in progenitors.

Figure 4: Cell type specific splicing quantitative loci (sQTL) analysis.

(A) A schematic illustrating splicing QTL mapping. Association of variants locating within 200 kb distance from each end of intron junctions were tested. The T allele is associated with more frequently splicing of the shorter intron junction.

(B) Differential splicing of two intron junctions within the *DLG4* gene. Splice 1 (chr17:7191358-7192945) supports a previously validated nonsense mediated decay transcript (ENST00000491753) with higher expression in progenitors, whereas splice 2 (chr17:7191358-7191893) has higher expression in neurons.

(C) Density plot showing the distribution of primary sSNPs relative to the two edges of intron junctions (progenitor in purple in upper panel, neuron in green in lower panel).

(D) Two intron junctions supporting an alternative 3' splicing site for *TMEM216* gene regulated by variant rs11382548 located at the splice site. The regional association of variants to two introns is shown in the genomic track at the left colored by pairwise LD r^2 relative to variant rs11382548. Box plots show quantile normalized PSI values for splice 1 (chr11:61397975-61398261) and splice 2 (chr11:61397975-61398270) across variant rs11382548.

(E) Enrichment of cell type specific sQTLs in functional genomic annotations at eigenMT-BH threshold used to define significant sQTLs. Each annotation is listed in the y-axis, and the x-axis shows the effect size of enrichment test, where data points are colored by -log10(p-value) from enrichment test and asterisks are used to denote significance (progenitor at the left, neuron at the right).

(F) Enrichment of cell type specific sQTLs within RNA-binding protein (RBP) sites based on a CLIP-seq dataset. The top 30 RBPs based on -log10(enrichment p-value) are listed on the y-axis, and the x-axis shows the effect size from enrichment test, where data points colored by -log10(p-value) from the enrichment test and cell type specific RBP are colored with purple for progenitors at the left, and as green for neuron at the right).

(G) Overlap percentage of cell type specific sSNP-intron junction pairs shared (pink) with fetal bulk sQTLs. Odds ratio test p-value is shown.

Figure 5: Colocalization of cell type specific eQTLs with GWAS for brain related traits.

(A) Number of GWAS loci colocalized with progenitor (purple), neuron (green) specific eQTLs or both cell types (orange). Each GWAS trait is listed on the y-axis.(SA: surface area, TH: thickness)

(B) Overlap of colocalized GWAS loci-gene pairs per trait combinations across progenitor, neuron and fetal bulk eQTL colocalizations for the traits listed in Figure 5A.

(C) Genomic track showing regional association of variants with educational attainment (EA), global surface area (GSA) and *CENPW* gene expression in progenitors and neurons. Progenitor eSNP rs4897179 (3rd row) was coincident with index SNP (rs9388490) for both EA (1st row) and GSA GWAS (2nd row), and conditioning progenitor eSNP rs4897179 on rs9388490 showed colocalization of the two signals (5th row). Also, rs4897179 was colocalized with another variant rs9388486 located in the chromatin accessibility peak at the promoter of *CENPW* gene (6th and 8th rows). Genomic tracks were color-coded based on LD r^2 relative to the variant rs9388486.

(D) Plot showing the chromatin accessibility peak (chr6:126339531-126340960) in progenitors across different genotypes of variant rs9388486, where allele C decreases chromatin accessibility (upper left). C allele of variant rs9388486 disrupted binding motifs of transcription factors including CREM, ATF1, ATF2 and ATF4 (bottom left). A schematic showing that potential transcription factor has decreased preference to bind at the C allele, which results in lower *CENPW* expression, increase in global surface area and educational attainment (right).

(E) Box plots showing chromatin accessibility across rs9388486 genotypes in progenitors (purple) and neurons (green) (upper panel). Box plots showing VST normalized *CENPW* gene expression across rs9388486 genotypes in progenitors (purple) and neurons (green) (bottom panel).

Figure 6: Colocalization of cell type specific sQTLs with GWAS for brain related traits.

(A) Number of GWAS loci colocalized with progenitor (purple), neuron (green) specific sQTLs or

both cell types (orange). Each GWAS trait is listed on the y-axis (SA: surface area, TH: thickness).

(B) Overlap of colocalized GWAS loci-intron junction pairs per trait across progenitor, neuron and fetal bulk sQTL colocalizations for the traits listed in Figure 6A.

(C) Genomic tracks showing regional association of variants with SCZ and an unannotated exon skipping splice site (chr10:102869593-102872448) for *AS3MT* gene in progenitors and neurons. The splice site was associated with progenitor sSNP (rs3740400) colocalized with SCZ GWAS index SNP (rs11191419). Genomic tracks were color-coded based on LD r^2 relative to the variant rs3740400. Sashimi plots showing the gene model of *AS3MT* and the genomic position of unannotated splice site (blue) overlapping with *AS3MT* gene. Average INT normalized PSI values for the splice site are shown for each genotype group. Schizophrenia risk allele T regulates the exon skipping event in progenitors. Boxplots showing INT normalized PSI values for splice across rs3740400 genotypes in progenitors and neurons.

(D) Genomic tracks color-coded based on pairwise LD r^2 relative to the variant rs1222218 showing regional association of variants with SCZ and an unannotated alternative splicing event for *ARL14EP* gene in progenitors and neurons. A cryptic exon skipping splice site (chr11:30323202-30332866) was associated with progenitor sSNP (rs1222218) colocalized with SCZ GWAS index SNP (rs1765142). Sashimi plots with the gene model of *ARL14EP* and the genomic position of the unannotated splice site (blue) overlapping with *ARL14EP* gene. Average INT normalized PSI values for the splice site are shown for each genotype group. Schizophrenia risk allele G regulates the exon skipping event in progenitors. Boxplots showing INT normalized PSI values for splice across rs1222218 genotypes in progenitors and neurons.

Figure 7: Prediction of gene expression during human brain development via TWAS.

(A) Manhattan plots for schizophrenia, IQ, and neuroticism TWAS for progenitors (purple-grey, top) and neurons (green-grey, bottom) where the LD matrix used was based on a European

population. Each dot shows -log10(TWAS p-value) for each gene on the y-axis, genes were color-coded based on discovery also in colocalization analysis (orange), defined as the nearest gene to GWAS loci (dark pink), being in both these two categories (blue), and discovered only in TWAS analysis (black). Only joint independent genes are labelled (positively and negatively correlated genes represented by triangle and square, respectively).

(B) IQ TWAS results for the *B3GALNT2* gene, regional association of variants to IQ trait shown at the top, and statistics from each TWAS study shown at the bottom (red line used for genome-wide significant threshold 5×10^{-8}).

Supplementary Figure 1: Study pipeline and immunostaining of progenitors and neurons

(A) Flow cytometry results showing sorting of EGFP positive neurons in pink.

(B) Immunostaining images indicates outer radial glia marker HOPX in green, proliferation marker Ki67 in yellow and pan-radial glia marker SOX2 in red were expressed in progenitors, and neuronal marker TUJ1 was expressed in neurons from 8 week differentiated cultures (scale bar is 100 μm, DAPI in blue).

(C) Computational pipeline used in our study: RNA-sequencing pipeline including steps for gene expression, allele specific expression and splicing at the left and genotyping pipeline that resulted in genetic mapping of different phenotypes derived from RNA-sequencing data.

Supplementary Figure 2: Quality control of RNA-seq data and covariate selection

(A) Replicate correlation of RNA-seq libraries across donors and within donors. Gene expression profiles were more correlated between libraries generated from the same donor thawed at different times as compared to libraries across different donors for both progenitors (left, p-value=0.001775) and neurons (right, p-value=0.0189).

(B) Principal component analysis (PCA) before and after batch correction of neuron for the machine (Sony SH800S in blue, FACS Aria II in red, progenitors not sorted in grey) used for sorting.

(C) Correlation of technical confounders with top 10 principal components of gene expression in progenitor, neurons and all data (asterisk indicates significant correlation).

(D) Correlation of technical confounders with top 10 principal components of splicing in progenitor, neurons and all data (asterisk indicates significant correlation).

(E) Multidimensional scaling (MDS) of global genotypes showing the multi-ancestry donors in our study. MDS1 vs MDS2 values plotted where each red circle represents a unique donor in our study and each different color represents different ancestry from HapMap3 (ASW: African ancestry, CEU:Northern and Western European ancestry, CHB: Han Chinese ancestry, CHD: Chinese in metropolitan Denver, GIH: Gujarati Indians in Houston, JPT: Japanese in Tokyo, LWK: Luhya in Webuye, MEX:Mexican ancestry, MKK: Maasai in Kinyawa, TSI: Toscani in Italy, YRI: Yoruba in Ibadan)

(F) Covariate selection analysis for eQTLs with number of eGene vs. number of global gene expression PCs (left, progenitors in purple, neurons in green), and for sQTLs with number of significant intron vs. number of global splicing PCs (right, progenitors in purple, neurons in green). Blue arrows indicate the number of PCs used in each dataset.

(G) Comparison of genomic inflation factor (λ_{GC}) without controlling for population structure and technical confounders (no control), only controlling for technical confounders by adding global gene expression PCs, controlling for both population structure (10 MDS of global genotype) and technical confounders, and controlling for kinship matrix in addition to the previous covariates.

Supplementary Figure 3:

(A) Comparison of the transitions between every two developmental points of in vivo developing cortex and *in vitro* progenitor and neurons with rank-rank hypergeometric overlap (RRHO) maps. The extent of overlap between in vivo and in vitro transcriptome was represented by each heatmap colored based on -log10 p value from a hypergeometric test with step size = 200. Each map shows the extent of overlapped upregulated genes in the bottom left corner, whereas shared downregulated genes are displayed in the top right corners.

Supplementary Figure 4: Conditional e/sQTL analysis

(A) A schematic showing the conditional e/sQTL procedure. Conditionally independent SNPs were found conditioning on the genetic variant with the most significant association, and iteratively applying the same algorithm until there were no further significant associations with local variants.

(B) Number of eGenes on the x-axis regulated by number of conditionally independent eSNPs on the y-axis indicated by eQTL order.

(C) Number of intron junctions on the x-axis regulated by number of conditionally independent sSNPs on the y-axis indicated by sQTL order.

Supplementary Figure 5: Cell type and temporal specificity of eQTLs

(A) Overlap between progenitor and neuron eQTLs for eSNP-eGene pairs and eGenes.

(B) Posterior probability of shared effect size (m-value). Upper pie shows the percentage of primary neuron eQTLs shared with progenitor eQTLs, and lower pie shows the percentage of primary progenitor eQTLs shared with neuron eQTLs at m-value > 0.9.

(C) Overlap between progenitor/neuron eQTLs and adult brain cortex eQTLs for eSNP-eGene pairs.

Supplementary Figure 6: Cell type specific allele specific expression

(A) MA plot showing allele specific expression for progenitor (purple) and neuron (green).

(B) Percentage of all ASE sites used in the analysis that overlaps with exonic regions.

Supplementary Figure 7: Cell type and temporal specificity of sQTLs

(A) Overlap of intron junctions, sGenes that are the genes intron junctions span and sSNP-intron junction pairs for progenitor vs neuron sQTLs

(B) Posterior probability of shared effect size (m-value). Upper pie shows the percentage of primary neuron sQTLs shared with progenitor sQTLs, and lower pie shows the percentage of primary progenitor sQTLs shared with neuron sQTLs at m-value > 0.9.

(C) Comparison of cell type specific sQTL vs eQTLs, progenitor in purple and neuron in green. Overlap between sGenes and eGenes, upper panel; overlap between sGene-sSNP and eGene-eSNP pairs, lower panel.

(D) Overlap between progenitor(in purple)/neuron (in green) sQTLs and adult brain cortex sQTLs (in red) for intron junction-sSNP pairs.

Supplementary Figure 8: Colocalization of cell type specific e/sQTLs with GWAS traits

(A) A schematic summarizing colocalization analysis including steps (1) to define coincident e/sQTLs with GWAS based on LD r^2 calculated either for European or study population, and (2) conditioning e/sQTLs on coincident GWAS index SNP.

Supplementary Figure 9: Prediction of gene expression/splicing during human brain development via TWAS.

(A) Manhattan plots for schizophrenia, IQ and neuroticism TWAS for progenitors (purple-grey, top) and neurons (green-grey, bottom) where LD matrix was calculated based on the population included in our QTL study. Each dots show -log10(TWAS p-value) for each gene on the y-axis, genes were color-coded based on being discovered also in colocalization analysis (orange), defined as the nearest gene to GWAS loci (dark pink), being in both these two categories (blue), and discovered only in TWAS analysis (black). Some of the jointly independent genes are labelled (positively and negatively correlated genes represented by triangle and square, respectively, full list is available in Table S8).

(B) Manhattan plots for schizophrenia, IQ and neuroticism TWAS for progenitors (purple-grey, top) and neurons (green-grey, bottom) where the LD matrix was calculated based on the population included in our QTL study. Each dots showing -log10(TWAS p-value) for each intron junctions on the y-axis, introns were color-coded based on discover also in colocalization analysis (orange), and being joint independent (asteriks)(positively and negatively correlated splicing represented by triangle and square, respectively).

(C) Comparison of TWAS genes and introns performed by using different LD matrices based on European (LD European) and population included in our QTL study (LD Study).

(D) Overlap of cell type specific TWAS genes/introns (from the analysis where LD was estimated from European population) with fetal brain bulk and adult brain bulk TWAS genes/introns.

Supplementary Figure 10: Prediction of alternative splicing events during human brain development via TWAS.

(A) Manhattan plots for schizophrenia, IQ and neuroticism TWAS for progenitors (purple-grey, top) and neurons (green-grey, bottom) where LD matrix calculated based on a European population. Each dot shows the -log10(TWAS p-value) for each intron junctions on the y-axis, introns were color-coded based on discovery also in colocalization analysis (orange), and being jointly independent (asteriks), where positively and negatively correlated splicing represented by triangle and square, respectively.

(B) SCZ TWAS results for intron junction (splice, chr7:2235564-2239418) of the *MRM2* gene, regional association of variants, that were used to test polygenic impact on introns to SCZ are shown on the left. Gene-model for *MRM2* is shown on the right with matching introns and statistics from each TWAS study shown at the bottom (red line used for genome-wide significant threshold of 5×10^{-8}).

Supplementary Tables

Table S1: Differential gene expression analysis progenitor vs neurons (FDR < 0.05): gene is the ensemblID, logFC is the expression fold change logFC > 0 indicates a gene more frequently expressed in neurons than progenitors; AveExpr is the average vst normalized expression of all samples. t is the expression fold change divided by its standard error[73]. P.Value is the nominal p-value from the testing differential expression; adj.P.Val is the Benjamini-Hochberg FDR adjusted p-value; B is log-odds for the differentially expressed gene in limma.

Table S2: List of cell type specific conditionally independent eQTLs for progenitor, neurons and fetal bulk: snp is the variant tested in QTL; beta is the beta coefficient; pval is the nominal p-value; gene is the ensemblID of the gene tested; rank is the eQTL order; chr is the chromosome number, BP is the genomic position of the variant; Cond.beta is the beta after conditional analysis; Cond.pval is the p-value after conditional analysis; A1 is the effect allele; rsid is the rs id of the allele matching in 1000 Genome Phase 3.

Table S3: Allele specific expression analysis (FDR < 0.05). SNP is the variant tested for allele specific expression analysis, baseMean is the average of the normalized count values divided by size factors from DESeq2[74]; log2FoldChange is the expression fold change logFC > 0 indicates reads more frequently expressed in donors with reference allele than donors with alternative allele; lfcSE is the standard error estimate for log2FoldChange; stat is the test statistics performed in DESEq2; pvalue is the nominal p-value from the testing differential expression; padj is the Benjamini-Hochberg FDR adjusted p-value; refAllele is the reference allele of the variant.

Table S4: Differential splicing analysis progenitor vs neurons (FDR < 0.05): intron is the splice junction, logFC is the expression fold change logFC > 0 indicates a gene more frequently expressed in neurons than progenitors; AveExpr is the average vst normalized expression of all samples. t is the expression fold change divided by its standard error[73]. P.Value is the nominal p-value from the testing differential expression; adj.P.Val is the Benjamini-Hochberg FDR adjusted p-value; B is log-odds for the differentially expressed intron in limma; chr is the chromosome, start is the start position of the junction; end is the end position of the junction; clusterID is the cluster identified from Leafcutter, cluster is the clusterID combined with chromosome number, verdict is the annotation status; gene is the gene symbol of the gene that introns junctions overlap with; ensemblID is the ensemblID of that gene; transcripts is the

transcripts where intron junction overlap with; constitutive.score: degree of the junction shown in each transcript.

Table S5: List of cell type specific conditionally independent sQTLs for progenitor, neuron and fetal bulk sQTLs: snp is the variant tested; beta is the beta coefficient, pval is the nominal p-value; intron is the intron junction as chromosome:start position:end position format; rank is the order of sQTL after conditional analysis; chr is the chromosome, start is the start position of the junction; end is the end position of the junction; clusterID is the cluster identified from Leafcutter, cluster is the clusterID combined with chromosome number, verdict is the annotation status; gene is the gene symbol of the gene that introns junctions overlap with; ensemblID is the ensemblID of that gene; transcripts is the transcripts where intron junction overlap with; constitutive.score: degree of the junction shown in each transcript; Cond.beta is the beta coefficient after conditional analysis (for primary QTLs, it is identical to beta); Cond.pval is the p-value after conditional analysis (for primary QTLs, it is identical to pval), A1 is the effect allele; rsid is the rs id of the allele matching in 1000 Genome Phase 3.

Table S6: Enrichment of RNA binding protein sites within cell type specific sQTLs. PThresh is the p-value threshold used for enrichment; OR is the odd ratio; Pvalue is enrichment p-value; Beta is the beta coefficient after enrichment test via GARFIELD[75]; SE is the standard error; CI95_lower is the lower bound of 95% confidence interval; CI95_upper is the upper bound of 95% confidence interval; NAnnotThesh is the is the number of annotated variants at the p-value threshold; NAnnot is the total number of variants after pruning; NThresh is the number of variant passing p-value threshold after pruning; N is the number of variants remained after pruning; linkID is the ID in annotation file; Annotation is the RNA-binding protein; Tissue is the the type of tissue used; Type is the cell type used for enrichment test.

Table S7: Colocalization of GWAS for neuropsychiatric disease and other brain related traits with cell type specific e/sQTLs and fetal bulk e/sQTLs: e/sQTLsnp is the e/sSNP; inibeta is the

beta coefficient before conditioning on GWAS SNP; pval is the nominal p-value prior to conditional analysis, gene/intron is the ensemblID of gene/intron junction associated with the e/sSNP; A1 is the effect allele for e/sQTL index SNP; GWASsnp is the variant e/sSNP colocalized with; Condbeta is the beta estimate of e/sQTL after conditional analysis; Condpval is the p-value after conditional analysis; r2 is the linkage disequilibrium (LD) r^2; pop is the population used to estimate LD r^2 (European population, with "European" or the population used in the QTL study with "Study"); symbol of the symbol of the gene (for eQTLs); biotype is the biotype of the gene for eQTLs; trait is the trait for GWAS.

Table S8: List of cell type specific/fetal bulk/adult bulk TWAS gene and introns for neuropsychiatric disease and other brain related traits. Output from FUSION[60,61]: ID is the gene ensemblID or intron id; CHR is the chromosome number; HSQ is the heritability; BEST.GWAS.ID is the GWAS SNP in the locus with the most significant association; BEST.GWAS.Z is the z-score of the best GWAS SNP; EQTL.ID is the best e/sQTL in the locus; EQTL.R2 is the cross-validation R^2 of the best e/sQTL in the locus; EQTL.Z is the z-score of the best e/sQTL in the locus; EQTL.GWAS.Z is the GWAS Z-score for this e/sQTL; NSNP is the number of SNPs in the locus; NWGT is the number of snps with non-zero weights; MODEL is the best performing model; MODELCV.R2 is the the cross-validation R^2 of the best performing model; MODELCV.PV is the p-value from the cross-validation of the best performing model; TWAS.Z is the TWAS z-score; TWAS.P is the TWAS p-value; trait is the GWAS trait; pop is the population used to estimate LD; joint_independent is the status if a gene/intron jointly independent (YES, if it is independent; NO, if it is not independent; NA, if it was not tested for the trait).

Table S9: Summary of heritability (p-value < 0.01) and cross validation r^2 from prediction models across cell type specific/fetal bulk/adult bulk for gene and intron TWAS: hsq is the mean

heritability of the genes/introns; hsq.se is the mean standard error of estimated heritability; hsq.pv is the mean p-value of the heritability; emmax.rsq is the mean cross-validation R^2 training via EMMAX with p-value as emmax.pval; lasso.rsq is mean the cross-validation R^2 via LASSO with p-value as lasso.pval; enet.rsq is mean the cross-validation R^2 via elastic net with p-value as enet.pval; blup.rsq is mean the cross-validation R^2 via BLUP with p-value as blup.pval; bslmm.rsq is the mean cross-validation R^2 via BSLMM with p-value as bslmm.pval; top1.rsq is the mean cross-validation R^2 via standard marginal eQTL Z-scores computation with p-value as top1.pval. 95 % confidence intervals per parameter are shown their below.

Methods

Cell Culture

Generation of human neural progenitor cells was previously described[16,26]. Briefly, human fetal brain tissue was acquired from the UCLA Gene and Cell Therapy Core following IRB regulations from approximately 14-21 gestation weeks (inferred to be 12-19 postconceptional weeks). Presumably cortical tissue was selected by visual inspection, subjected to single cell dissociation, and cultured as neurospheres. Neurospheres were plated on laminin/fibronectin and polyornithine coated plates for an average of 2.5 ± 1.8 s.d. passages, and cryopreserved as primary human neural progenitors (phNPCs).

Cryopreserved phNPCs were transferred to UNC Chapel Hill, after material transfer agreement, where all downstream culture and analyses were completed. Donors processed for ATAC-seq (described previously[16]) and RNA-seq (described here) were cultured simultaneously. The overall design of the experiment and media used for culture was previously described[16]. Briefly, we cultured 89 unique donors for subsequent RNA-seq library preparation. We first randomly assigned the approximately 8-9 donors into 10-12 rounds for a feasible cell culture workload. We thawed one round every three weeks. To reduce batch effects, we processed each round

on the same day of the week and designated the same person to do each task as much as possible. Cells were isolated at two time points: progenitor and their differentiated and virally labeled neuronal progeny. Progenitors were cultured in proliferation media including growth factors for 3 weeks, and we lifted them with trypsin to prepare RNA-seq libraries. Differentiation was performed for 5 weeks, after which the culture was transduced with AAV2-hSyn1-eGFP (https://www.addgene.org/50465/) virus at 20,000 multiplicity of infection (MOI) to label neurons and then differentiated for another 3 weeks. FACS sorting (using either BD FACS Aria II or Sony SH800S) at 84 days post-differentiation was used to isolate EGFP labelled neurons (Supplementary Figure 1A). After cells were isolated as either progenitors or neurons, we added Qiazol and stored the mixture at -80°C for randomized RNA isolation to reduce batch effects.

Immunofluorescence labeling and imaging

At the progenitor stage or after 8 weeks of differentiation, we fixed the cells by incubating them in 4% PFA, and performed permeabilization with 0.4% Triton in PBST. We used 10% goat serum dissolved in PBST for blocking. We incubated blocked samples with primary antibodies dissolved in PBST solution with 3% goat serum at 4°C overnight followed by washing 3 times with PBST. Samples were subject to incubation in fluorophore-conjugated secondary antibodies, for 1 hour at room temperature, then they were stained with DNA-binding dye DAPI with 10 minutes incubation. We used antibodies with concentrations listed as following: SOX2 (1:400, rabbit, Millipore #AB5603), Ki67 (1:1000, rat, Invitrogen #14-5698-82), TUJ1 (1:2000, mouse, Biolegend #801202), Alexa Fluor 568 (1:1000, goat anti-rabbit, Invitrogen #A11036), Alexa Fluor 647 (1:1000, goat anti-rat, Invitrogen #A21247), Alexa Fluor 488 (1:1000, goat anti-mouse, Invitrogen #A11001).

RNA-seq Library Preparation

We isolated RNA from progenitors and neurons using Qiagen miRNeasy Minelute kit, quantified RNA concentration with a Qubit 2.0 fluorometer, and assessed RNA integrity via eRIN scores using the Agilent Tapestation. We prepared libraries for sequencing using Kapa Biosystems KAPA Stranded RNA-seq with Riboerase (HMR) kit by loading 50 ng of total RNA into the initial reaction. We followed the manufacturer's instructions for fragmentation and PCR steps. To obtain ~350 bp average insert size, we fragmented cDNA at 85°C for 6 min. Final library concentrations were determined using Qubit 2.0 fluorometer and pooled to a normalized input library. Pools were sequenced on a NovaSeq S2 flowcell using 150 bp PE reads with an average read depth of 99M per sample.

RNA-sequencing data processing

We merged fastq files from the same library when sequenced on multiple flow cells and trimmed the adapters using sequences provided by Illumina with Cutadapt/1.15[76]. Quality control of each library was performed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For alignment, we first integrated the sequence of AAV2-hSyn1-eGFP plasmid (https://www.addgene.org/50465/) used for labeling neurons into GRCh38 release92 reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.38/). Then, we aligned the fastq files to this combined reference genome by implementing STAR/2.6.0a aligner[77].

We processed aligned data further with different steps based on downstream analyses (Supplementary Figure 1C). To estimate gene expression levels, we quantified reads with the union exon based approach using featureCounts, where for each gene, all overlapping exons were merged to form union exons, and the reads mapped to those union exons with the same strandedness were counted[78]. Gene models were identified using the GTF file Homo_sapiens.GRCh38.92

(http://ftp.ensembl.org/pub/release-92/gtf/homo_sapiens/Homo_sapiens.GRCh38.92.gtf.gz) merged with AAV2-hSyn1-eGFP plasmid.

For allele specific expression and splicing quantification, we remapped the aligned data with WASP software(v2018-07)[79] to reduce reference mapping bias. First, we identified reads overlapping with bi-allelic SNPs within our acquired genotype data. Following this, the genotype of any reads overlapping with a SNP was swapped with the other allele, and re-mapped. WASP discarded re-mapped reads that did not map to the same genomic position. As a final step, we implemented *rmdup.py* script provided in the WASP software which removes duplicate reads randomly, regardless of their mapping score.

Mycoplasma contamination test

Adapter trimmed reads (see above) were mapped using STAR to a combined reference including the GRCh38 release 92 human reference genome, AAV2-hSyn1-eGFP plasmid, and over 1400 mycoplasma genomes. Alignment parameters allowed for simultaneous mapping of reads to one or more human and mycoplasma genomes. No sample exceeded 0.11% of total reads mapping to any mycoplasma genome, indicating none of our cultures were contaminated with mycoplasma. This mapping was only used for mycoplasma contamination analysis and not for subsequent analyses.

Genotype processing

We performed genotyping using Illumina HumanOmni2.5 or HumanOmni2.5Exome platform, and exported SNP genotypes to PLINK format following the procedure previously described[16]. Briefly, we converted SNP marker names from Illumina KGP IDs to rsIDs using the conversion file provided by Illumina. We performed quality control with PLINK v1.90b3 software[80] (Supplemental Figure 1C) as follows. We filtered out SNPs with the following criteria: variant

missing genotype rate > 5% $(--geno\ 0.05)$, deviations from Hardy-Weinberg equilibrium at $p<1x10^{-6}$ $(--hwe\ 1x10^{-6})$, minor allele frequency < 1% $(--maf\ 0.01)$. We also filtered out individuals with missing genotype rate > 10% $(--mind\ 0.10)$. We obtained 1,760,704 directly genotyped variants surviving our QC procedure. Lastly, we called sex from genotype data using PLINK v1.90b3 software based on heterozygosity on the X chromosome. When there was an ambiguity for sex assessment based on genotype data, we checked Xist gene expression. We estimated the population structure of our study cohort by implementing multidimensional scaling (MDS) for genotype data of our samples and genotype data from HapMap3 (https://www.sanger.ac.uk/resources/downloads/human/hapmap3.html). We followed the protocol from ENIGMA consortium (http://enigma.ini.usc.edu/wp-content/uploads/2012/07/ENIGMA2_1KGP_cookbook_v3.pdf). By plotting MDS1 vs MDS2, we visually show each donor's ancestry relative to known populations (Supplemental Figure 2E).

Imputation

After filtering genotype data, we pre-phased the data with SHAPEIT v2.837[81]. For our imputation reference panel, we used 1000 Genomes Project Phase 3 that contains a total of 37.9 million SNPs in 2,504 individuals with multiple ancestries, including those from West Africa, East Asia and Europe[82]). Imputation was implemented using Minimac4 software[83] (v1.0.0). On the X chromosome, we separately performed pre-phasing and imputation steps for the pseudoautosomal region and non-pseudoautosomal regions. Following imputation, we retained any variants with missing genotype rate lower than 0.05, Hardy-Weinberg equilibrium p value lower than 1×10^{-6} and minor allele frequency (MAF) bigger than 1%. We retained SNPs with sufficient imputation quality ($R^2 > 0.3$), and obtained approximately 13.6 million SNPs in total.

Sample quality control

One library with missing eRIN score and one library with missing final cDNA concentration from neurons were removed. In order to detect sample swaps or mixing between samples, we evaluated consistency of genotypes called from the RNA-seq and genotyping array via VerifyBamID v1.1.3[84]. We removed RNA-seq libraries file with [FREEMIX] > 0.04 or [CHIPMIX] > 0.04 ($N_{library}$ = 8). Also, we corrected samples where we detected swaps ($N_{library}$ = 7). After quality control, we retained 85 unique donors for progenitors, and 74 unique donors for neurons for subsequent analyses.

Replicate correlation and determination of technical factors correlating with gene expression

Quantified RNA-seq reads with featureCounts were imported to generate a gene count matrix in DESeqDataSet format from DESeq2 R package[74]. We filtered out the lowly expressed genes (those where fewer than 10 read counts of a gene were observed in fewer than 5% of samples), and normalized the data via variance stabilizing transformation (vst()) function from DESeq2 R package[74]. We included genes on the X and Y chromosomes and genes transcribed from mitochondrial DNA meeting the expression criteria. We subset the normalized gene expression matrix into progenitor and neuron specific samples. To identify major axes of variation in gene expression across samples, we computed principal components of gene expression with prcomp() function from stats R package for each cell-type separately, and reported the proportion of variance explained by each component.

We recorded biological and technical variables for each sample which may potentially impact gene expression: cell type, postconception week, sex, tissue acquisition date, researcher extracting RNA and preparing libraries, RNA input amount, index number and bases, final cDNA concentration, BioAnalyzer run date, average basepair of BioAnalyzer cDNA, sequencing

pool, cell input, Qiazol lot number and addition date, eRIN, RNA extraction date, RNA tapestation date, Qiagen extraction kit lot number, FACS sorting date and time, total live cells during sorting, FACS machines used, researcher performing FACS sorting, papain lot number and addition date, differentiation rank (a qualitative assessment of cell health evaluated under the microscope), well location in the 6-well plate, date to plate for differentiation, researcher washing and differentiating cells and date, virus addition date, researcher adding virus, PBS lot number used for cell proliferation and differentiation, laminin, polyornithine lot numbers used for proliferation and differentiation, donor ID, round, media lot numbers used for proliferation, passage number, split dates, researcher performing each split, rank for proliferation (qualitative assessment of cell health), trypsin lot number used for splitting cells, and fibronectin lot number. To identify technical covariates impacting expression levels, we assessed if any recorded biological or technical variables were significantly correlating with first 10 expression PCs separately for each cell type. We observed that different FACS machines (Sony SH800S with N_{donor} = 8; FACS Aria II with N_{donor}= 66) used to isolate GFP labelled neurons had a strong impact on global gene expression in neurons (PC1: r = 0.59, p-value = 1.782e-08; PC2: r = 0.58, p-value = 3.972e-08) (Supplementary Figure 2C). To remove the impact of sorter on global neuron expression profiles prior to differential expression analysis, we implemented $limma::removeBatchEffect$ function[73]. Then, we combined the gene expression matrix from batch corrected neurons with progenitors gene expression data.

We cultured several donors multiple times during the course of the experiment in order to quantify cell culture induced noise. We calculated Pearson's correlation of gene expression between libraries from the same donors ($N_{correlation}$ = 11 in both progenitors and neurons), and between each library across donors in a pairwise manner ($N_{correlation}$ = 11,556 for progenitors; $N_{correlation}$ = 9,312 for neurons). For neurons, we used gene expression values after batch correction with the limma R package for the sorter type, as described above. We performed an

unpaired two-sided t-test for statistical assessment of mean difference between these two categories after fisher's z transformation of correlation r values.

Differential gene expression analysis

We identified differentially expressed genes between progenitors and neurons by using vst normalized expression values corrected for sorter with limma R package[73]. To perform a paired differential gene expression analysis which inherently controls for donor related differences, we established the following design matrix: $model.matrix(\sim CellType + as.factor(DonorID) + RIN, data)$. Following this, we adjusted p values for each gene via multiple test correction with the Benjamini-Hochberg procedure[85], and defined significant differentially expressed genes as adjusted p value < 0.05.

Gene Ontology analysis

We performed gene ontology enrichment analysis by using the gprofiler2 package as the R interface to the g:Profiler tools by using GO:BP database[86]. For differentially expressed genes, after performing DGE analysis, we categorized the genes into two groups as upregulated in progenitors (logFC < -1.5 and adjusted p value < 0.05), and upregulated in neurons (logFC > 1.5 and adjusted p value < 0.05) (Figure 1E). For each enrichment analysis, we applied multiple correction test, and considered only pathway enrichments with adjusted p value lower than 5% false discovery rate as statistically significant.

Transition Mapping (TMAP)

To evaluate the transcriptomic similarity between our *in vitro* culture system and the *in vivo* brain, we performed transition mapping analysis as described in our previous work[26,87]. To evaluate transcriptomic similarity to cortical laminae in the developing brain, we used previously published laminar expression data from laser capture microdissected of prenatal human brain[32]

(H376.IIIB.02. female, 16 pcw, brainspan.org). In our comparison, genes were retained which showed expression in either cell-type and were present on the array in which the *in vivo* data was acquired. We used gene symbols to find ensemblIDs, and used ensemblIDs to match with *in vitro* data. When multiple probes were present for a given gene on the array, the probe with the highest expression per gene was used. We quantile normalized the gene expression, and we performed in vivo differential gene expression via limma between every two laminae. Similarly, we performed differential expression analysis in our *in vitro* cultures as described above. We applied transition mapping via RRHO2 R package with "stratified approach" to avoid misinterpretation of the discordant overlaps[88]. In this algorithm, firstly genes were ranked based on their degree of differential expression (DDE) (i.e., $-log10(p-value) \times signed\ effect\ size$) separately for *in vivo* and *in vitro* data. Following ranking, a hypergeometric test was applied to assess enrichment for each overlap between two datasets for a series of arbitrary cutoffs set through the highest degree to the lowest. By employing a stratified algorithm, we computed the degree of overlap. Finally, we visualized hypergeometric test -log10(p values) as a heatmap (Figure 1F).

To compare transcriptomic overlap between our *in vitro* system and developmental time points of the human brain, we downloaded gene expression data from a spatiotemporal atlas of the human brain[15,33], RNA-Seq Gencode v3c summarized to genes, brainspan.org). Only expression data for cortical regions were included, and the intersection of genes between *in vitro* with *in vivo* was defined as a gene list for subsequent analyses. We applied the same transition mapping procedure described above. Specifically, we estimated gene expression differences between developmental periods described in[33]. Because there were multiple samples from the same donor for each period for the *in vivo* data, we added donor id as a random effect within the design matrix for DGE analysis. Ranked genes by their DDE were

subjected to the RRHO test, and -log10 based p-values were converted into a heatmap plot (Supplementary Figure 3A).

Cell type specific local eQTL mapping

To perform local eQTL analysis, we conducted an association test between gene expression (retaining genes if at least 10 counts of the gene were present in more than 5% of the samples of that cell type, resulting in 24,767 and 27,638 genes for progenitors and neurons, respectively) with genetic variants within ± 1 Mb window of gene TSS for both autosomal chromosomes and X chromosome, for progenitors and neurons separately. Each gene TSS was defined as the transcription start site of the gene isoform with most upstream exon based on GTF file Homo_sapiens.GRCh38.92 (http://ftp.ensembl.org/pub/release-92/gtf/homo_sapiens/Homo_sapiens.GRCh38.92.gtf.gz).

We removed variants of low allele frequency in order to prevent one donor from strongly influencing association results. For variant selection, PLINK v1.90b3 software $--freqx$ function was implemented to obtain donor counts per genotype group for each variant. We included only variants with at least 2 heterozygous donors and no homozygous minor allele donors, or at least 2 minor allele homozygous donors for autosomal chromosomes, and for X chromosome we retained the variants with at least 2 haploid allele counts in addition to this criteria.

For eQTL mapping, we established a linear mixed effects regression model to control for population stratification and cryptic relatedness with EMMAX software[89]. To compute the kinship matrix, we implemented $emmax-kin\ -v\ -h\ -s\ -d\ 10$ algorithm creating the identity by state (IBS) kinship matrix by excluding all genetic variants located on the same chromosome as the tested variant from non-imputed genotype data for each single variant association test (MLMe

method; see [90]). We used additional ancestry control by including the first 10 MDS components from genotype data[91]. In order to control for unmeasured technical variables impacting gene expression, we sequentially added gene expression PCs and re-ran the genetic associations via EMMAX. For neurons, we included a covariate for FACS sorter for each run given its strong impact on gene expression.

The full association model for neurons was:

$$expression \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + FACS\ sorter + PCs\ of\ global\ gene\ expression$$

The full model for progenitors was:

$$expression \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + PCs\ of\ global\ gene\ expression$$

For each run, we adjusted nominal values of all gene variant associations, and defined significant associations with nominal p value lower than 5% false discovery rate (FDR)[85]. We found that 10 PCs and 12 PCs of gene expression resulted in a maximum number of eGenes discovery in progenitors and neurons, respectively (Supplementary Figure 2F). Our final eQTL model was:

Neuron:

$$expression \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + FACS\ sorter + 12\ PCs\ of\ global\ gene\ expression$$

Progenitors:

$$expression \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + 10\ PCs\ of\ global\ gene\ expression$$

In order to stringently control our association results for both number of variants and genes tested, we further implemented a hierarchical correction procedure called eigenMT-BH[37]. Using this method, as Step 1, we adjusted the nominal p values of the all cis SNPs separately for each gene to compute locally adjusted p values with eigenMT method[36] In Step 2, locally adjusted minimum p values for all genes were then subjected to BH procedure to obtain globally adjusted p values. In Step 3, we defined eGenes as genes with globally adjusted p value lower than 0.05. Then, to find other independent SNPs for those eGenes, we set the significance

threshold as the maximum nominal p value from step 1 that had corresponding globally adjusted p value lower than 0.05.

We performed conditional analysis by using this threshold p-value gathered from the eigenMT-BH multiple correction method to identify independent significant eQTLs. To identify conditionally independent eQTLs, for each eGene (a gene significantly associated with at least one variant), we iteratively included the hard call genotype of the variant with strongest association with eGene as a covariate, and re-ran the regression model specified above (Supplementary Figure 4A). We defined a variant as "conditionally independent" from the variant conditioned on, if the association of the variant with the eGene was still significant based on the initial threshold p-value. Then, we conditioned on those variants that met threshold p value condition at the first round plus the primary variant, and identified third conditionally independent eQTLs. We applied this procedure iteratively until no additional significant eQTLs remained[92,93].

Comparison of degree of controlling population structure between EMMAX vs FastQTL

We applied FastQTL[94] method in nominal pass mode for different models (1) without controlling for either for population structure or technical confounders (2) controlling for only technical confounders, (3) controlling for 10 MDS of global genotype and global gene expression PCs. Following this analysis, we compared genomic inflation factors (λ_{GC}) across those three groups to our data where we controlled for 10 MDS of global genotype and global gene expression, as well as the cryptic relatedness with kinship matrix.

Bulk Fetal brain eQTL mapping

We utilized bulk fetal cortical wall eQTL data described previously[19]. We re-analyzed data in this study with the following modifications to harmonize with the eQTL approach implemented in this study: (1) we controlled for population stratification using a linear mixed effects model as

described above, and (2) we included 23 additional donors which were genotyped after the publication of the previous manuscript. We used rRNA-depleted RNA-seq data from flash frozen human fetal brain cortical wall tissues derived from 240 donors at 14-21 gestation weeks (inferred to be 12-19 post conception weeks). We excluded 4 donors for sample swap and contamination based on verifyBAMID analysis, and one donor with sex ambiguity, resulted in 235 unique donors for eQTL analysis (35 of unique donors shared with cell type specific data). Gene based annotations of the genome were derived from Homo sapiens gene ensembl version 92 (GRCh38) for eQTLs. We included only genes with at least 10 counts in 5% of donors. We normalized the data with the VST method to be used as phenotype in eQTL analysis. We also extracted genomic DNA from the same donors, and performed genotyping on a dense array (Illumina Omni 2.5+Exome) and imputation to a common reference panel (1000 Genomes Phase 3; described above). Variants were retained in the analysis if there were at least 2 heterozygous donors and no homozygous minor allele donors, or if there were at least 2 minor allele homozygous donors as for cell type specific eQTLs, as described above.

We performed local eQTL analysis to test the association between each gene's expression and variants within the ±1 Mb window of the transcription start site of each gene. We applied linear mixed model association software, EMMAX[89] to control for population stratification and cryptic relatedness (as described above for cell type specific eQTL analysis). We used the linear mixed effects regression model testing association between expression of each gene and nearby genetic variants, controlling for 10 MDS genotype components, 10 PCs of gene expression, and a kinship matrix as random effect excluding the chromosome genotypes testing with the MLMe approach[90] . After association, nominal p-values were corrected for hierarchical multiple testing using the eigenMT-BH method as described above, and we obtained independent eQTLs performing conditional analysis as described for cell type specific eQTLs above.

Enrichment of eQTLs and sQTLs within functional genomic annotations

To identify enrichment of eQTLs and sQTLs within functionally annotated genomic regions, we implemented GARFIELD software to control for the distance to TSS, LD and minor allele frequency (MAF) of QTLs[75]. We used functional genomic annotations from 25 chromatin states given in the ChromHMM BED files of Roadmap Epigenomics project from human male fetal brain[41,95] lifted over from hg19 to hg38, splice sites, introns, 5' and 3' UTR. We extracted 5' and 3' splice sites (splice donor and splice acceptor sites) from Homo_sapiens.GRCh38.92 GTF file implementing hisat2_extract_splice_sites.py algorithm of hisat2 software (Kim, Langmead, and Salzberg 2015), and defined exon-intron boundaries as the region between those splice sites. 5' and 3' UTR were also defined based on the coordinates in the same GTF file. For all e/sQTLs, we extracted the p-value from the strongest association for each variant (with minimum p-value) in the case that one variant was associated with multiple genes/intron junctions. To create annotations files, we considered a variant overlapping with a functional element if the variant itself or any of the variants in high LD within 500kb ($r^2 > 0.8$) overlapped with each of annotation categories. LD pruning[75] was performed at $r^2 > 0.01$ within GARFIELD software. Following this, a logistic regression model controlling for the distance to TSS, LD proxies and MAF binned for five quantiles was performed with GARFIELD software for enrichment at eigenMT-BH p value thresholds defined in e/sQTL analysis. The effective number of annotations were estimated and multiple testing adjusted p values were computed by the software to identify enrichment of e/sQTLs within defined annotations.

RNA binding protein motif analysis

We performed sQTL enrichment in RNA binding protein binding sites via GARFIELD as described above. In this analysis, we used BED files including RNA binding protein sites from a

CLIP-seq database as annotation files[52], and assessed significant enrichment of cell type specific sQTLs for binding sites of each RBP.

Enrichment of cell type chromatin accessibility QTLs and GWAS SNPs within eQTLs

We applied QTLEnrich v2[23] to find enrichment of cell type specific caQTLs[16] within cell type specific, fetal bulk[119] and adult bulk eQTLs[23]. We extracted only variants within the chromatin accessibility peaks from caQTL data, and used the variant with the smallest p-value if the variant associated with open chromatin regions. We estimated LD using the study population used for our QTL study, and performed enrichment with 0.05 p-value cut off.

Allele specific expression analysis pipeline

To identify sites with Allele Specific Expression (ASE), we applied the ASEReadCounter algorithm from GATK tools[96] to the RNA-seq data remapped with WASP to reduce mapping bias and to discard duplicate reads. For each donor, we counted allele specific reads overlapping with bi-allelic variants identified in the genotypeVCF files. We retained only variants with at least 5 heterozygous donors, and at least 10 counts from either allele. ASE can be falsely called when genotyping errors are present in the dataset. We used two approaches to identify and remove potential genotyping errors: (1) We detected wrongly called variant genotypes by assessing concordance between genotypes called by DNA versus RNA[42]. We removed variants that were called homozygous based on the genotype data when at least 10 counts of the alternate allele were present in the RNA-seq data, (2) we discarded variants where at least 7 heterozygous donors based on genotype data have zero counts for one of the alleles, which may indicate a donor falsely called as heterozygote when in truth the donor is a homozygote (given that (1/2)^7 = 0.008, meaning that probability of having all donors receiving an imprinted allele from either mother or father is low) . Because ASEReadCount does not disambiguate the strandedness of reads, it is not possible to confidently assign reads

overlapping with multiple gene annotations to a specific gene[78]. Therefore, if a variant overlapped with more than one gene annotation, we removed the variant by implementing findOverlaps function from IRanges R package[97] for genes based on their genomic coordinates defined GTF file Homo_sapiens.GRCh38.92 (http://ftp.ensembl.org/pub/release-92/gtf/homo_sapiens/Homo_sapiens.GRCh38.92.gtf.gz).

To evaluate allelic imbalance, we used DESeq2 with the design:

$design = \sim 0 + RNAid + Allele$. Excluding homozygous donors, we computed the log2 fold change of non-reference allele counts over reference allele counts and used a Wald test to detect allelic imbalance. Multiple test correction was performed with the Benjamini and Hochberg method, and we defined significant ASE sites as those with adjusted p-values lower than 0.05.

To compare eQTLs with ASE sites (Figure 3C-D), we extracted eQTLs associations with the same filtering criteria of variants used for ASE analysis (at least 5 heterozygous donors and overlapping with at least 10 RNA-seq reads). We also extracted eGenes (defined based on significant eigenMT-BH global p value) with at least 10 counts per donor.

Quantification of Intron Excisions

To identify alternatively excised introns, separately for each cell type, we extracted exon-exon junctions from WASP-mapped RNA-seq data in BAM format via regtools $junctions\ extract$ function where reads map to a minimum of 6 nt of each exon[98]. Next, we processed those junctions that are called *intron excisions* or *exon-exon junctions* with the pipeline provided by Leafcutter software[48]. Firstly, intron excisions with shared splice junctions were clustered together applying an iterative procedure until each cluster has at least 50 reads across donors and introns with maximum 50 kb length, separately for progenitors and neurons. For differential splicing analysis, we performed clustering by combining exon-exon junctions files from each cell type. For each cluster, intron excisions supported by at least one count in more than 5 donors

(within each set of donors contributing to the 3 different sQTL analyses for that cell type (progenitor, neuron) or tissue class (fetal brain bulk); or for differential splicing analysis across donors from both cell types used (progenitor + neuron) were retained. We further calculated intron excision ratios, and filtered out introns represented in less than 40% of donors (within each set of donors contributing to the 3 different sQTL analyses for that cell type (progenitor, neuron) or tissue class (fetal brain bulk); or for differential splicing analysis across donors from both cell types used (progenitor + neuron) with $prepare_phenotype_table.py$ algorithm. We referred to each intron excision ratio as *percent spliced in* (PSI) that corresponds to the usage of each intron compared to other introns in the same cluster. Standardized and quantile normalized intron excision ratios, and global alternative splicing PCs computed with those ratios were used for downstream analysis.

Differential splicing analysis

To perform differential splicing analysis, we used quantile normalized PSI values as input to the limma package[73]. Identical to differential expression analysis, neuron splice ratios were corrected for batch including FACS machine used for sorting with $limma :: removeBatchEffect$ function. Batch corrected neuron splice ratios were combined with progenitor data. We implemented a paired differential splicing analysis inherently controlling donor related differences with design matrix: $model.matrix(\sim CellType + as.factor(DonorID) + RIN, data)$. We defined intron junctions with adjusted p values from via multiple test correction with Benjamini-Hochberg procedure[85] lower than 0.05 as significant differentially spliced introns.

Splicing QTL mapping

We performed cell type specific splicing QTL analysis by testing the association of PSI with the genetic variants located within ± 200 kb window from starting and end points of the splice junctions for autosomal chromosomes and the X chromosome. Identical to local eQTL analysis,

we used only genetic variants that met the following criteria: if there were at least 2 heterozygous donors and no homozygous minor allele donors, or if there were at least 2 minor allele homozygous donors. Furthermore, we controlled for population stratification and cryptic relatedness as described above for local eQTL mapping.

We used standardized and normalized intron excision ratios (percent spliced in) calculated by leafcutter as the phenotype for sQTL mapping. EMMAX[89] was used to test for association between SNPs within a cis-region of +/- 200kb of the intron cluster and intron ratios within cluster. We controlled for population stratification and cryptic relatedness as described above for eQTL mapping. Also, we controlled for unmeasured technical variables impacting alternative splicing by sequentially adding global splicing PCs to the genetic associations via EMMAX. Again for neurons, we additionally controlled for FACS sorter for each run given its strong impact on splicing as well.

The full model for neurons was:

$$PSI \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + FACS\ sorter + PCs\ of\ global\ splicing$$

The full model for progenitors was:

$$PSI \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + PCs\ of\ global\ splicing$$

For every run, we adjusted nominal values of all PSI variant associations, and defined significant associations with lower than at 5% false discovery rate (FDR)[85]. We found that 1 PC and 1 PC across the PSI matrix resulted in a maximum number of intron excisions with at least one significant association in progenitors and neurons, respectively (Supplementary Figure 2F). Our final sQTL model was:

Neuron:

$$PSI \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + FACS\ sorter + 1\ PCs\ of\ global\ splicing$$

Progenitors:

$$PSI \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + 1\ PCs\ of\ global\ splicing$$

Implementing a hierarchical correction procedure called eigenMT-BH[37], as step (1), we adjusted the p values of the all cis SNPs strongest association separately for each intron excision to compute locally adjusted p values with the eigenMT method[36], and then locally adjusted minimum p values for all intron excisions were subjected to the BH procedure giving globally adjusted p values. Intron excision with corresponding global p value lower than 0.05 were considered as significant alternative splicing events. In order to find other independent significant sQTLs in addition to the ones associated with lowest p values, we applied conditional analysis at eigenMT-BH p-value threshold as described for eQTL analysis.

For bulk fetal cortical tissue sQTL mapping, we applied the same strategy used for cell type specific sQTLs, and found the following model maximized significant intron junctions discovery:

$$PSI \sim SNP + 10\ MDS\ of\ global\ genotype + kinship\ matrix + 5\ PCs\ of\ global\ splicing$$

After calculating eigenMT-BH threshold p value, we performed conditional analysis to define independent significant sQTLs.

To find genes overlapping with intron excision, we annotated intron junctions by using Leafcutter based on genomic coordinates and gene model provided in GTF file Homo_sapiens.GRCh38.92. Intron junctions assigned as cryptic 5', cryptic 3', novel annotated pair were considered as novel splicing events for the genes overlapped with junctions. For unannotated splice sites for *AS3MT* and *ARL14EP* genes, we additionally checked for overlap with known splice sites up to Ensembl Release 100.

Estimation of m-values for cross cell comparison

We estimated m-values to assess cell type specificity of SNP-gene or SNP-intron excision pairs with Metasoft[99]. Prior to software implementation, we extracted e/sQTLs from the neuron data corresponding to primary progenitor e/sQTLs to see overlap of sharing significant progenitor

e/sQTLs with neuron eQTLs. Similarly, we extracted e/sQTLs from the progenitor data corresponding to primary neuron e/sQTLs to see overlap of sharing significant neuron eQTLs with progenitor e/sQTLs. We estimated standard errors via dividing beta estimates from EMMAX by t-statistics for each association p value. We considered associations shared across different QTLs for the m-value > 0.9.

Comparison of cell type specific vs fetal and adult bulk brain e/sQTLs

We considered an overlap of e/sQTLs between two datasets when the index e/sSNPs were in LD (r^2>0.8 where LD was calculated in our sample population) and the eSNP-eGene/sSNP-intron pairs were shared. To determine the total number of eSNP-eGene/sSNP-intron pairs as the universe for enrichment analyses, we pruned all variants associated with each gene per gene for r^2 > 0.01 by using PLINK command $plink -- indep - pairwise\ 50\ 5\ 0.01$. To determine if different proportions of sharing were observed between two cell types, we performed an odds ratio test described here[100].

To test temporal specificity of cell type specific e/sQTL data, we downloaded GTEx data adult brain e/sQTL data[23]. We called loci from the two datasets as colocalized when, (1) index adult brain e/sQTLs are found within LD buddies of cell type specific e/sQTLs at LD r^2 > 0.8 (where LD is calculated using either the European population from 1000 Genomes or our study's population), and (2) when the cell-type specific e/sQTL data conditioned on index adult brain e/sQTLs, the index brain e/sQTL no longer survives the global significance threshold.

LD-thresholded colocalization with brain disorders and traits GWAS

To find eQTLs and sQTLs colocalized with index GWAS loci, we performed LD-thresholded colocalization analysis for each cell type separately[55]. We used summary statistics of GWAS for schizophrenia (SCZ)[1], major depression disorder (MDD)[101], bipolar disorder (BP)[2], educational

attainment (EA)[102], neuroticism[103], IQ[6], cognitive performance (CP)[102], attention-deficit/hyperactivity disorder (ADHD)[104], Alzheimer's disease (AD)[105], Parkinson's disease (PD)[106], Insomnia[107], epilepsy[108], autism spectrum disorder (ASD)[109], and cortical thickness and surface area from ENIGMA project[5]. We liftovered positions of variants in GWAS summary statistics from hg19 to hg38 with $liftOver()$ function from R rtracklayer package[110]. Variant rsids were assigned with dbSNP151 based on positions of variants in summary statistics data. To define index GWAS SNPs at genome-wide significant threshold p value ($5x10^{-8}$), we implemented a clumping procedure, where we defined two LD-independent GWAS signals so as to have pairwise LD $r^2 < 0.5$ based on LD matrix computed with European population of 1000 Genomes (1000G European phase 3). Prior to clumping, duplicated rsIDs in 1000G EUR genotype files were assigned with unique names, and BIM files were modified for each chromosome. Following a unique id assignment, BIM files were merged back to BED and FAM files with --bmerge function of PLINK1.9 software (plink --bfile BED file --bmerge modified_BIM file). Since all GWAS we leveraged in our colocalization analysis have been conducted in populations of European ancestry, and our study population is multi-ancestry, we computed LD r^2 separately within these two different populations. We considered the index eQTL or sQTL SNP coincident with the index GWAS SNP if the pairwise LD r^2 between them was greater than 0.8 based on either the LD matrix computed via either European 1000 Genomes Phase 3 data or our study population. Following that, we performed a conditional eQTL/sQTL analysis by conditioning on the coincident index GWAS SNP. If the association of index QTL and gene expression or intron excision was no longer significant based on p value thresholds defined with eigenMT-BH method for each dataset, we identified that cell type specific and fetal bulk eQTL/sQTL as a colocalized loci with the given GWAS trait (Supplementary Figure 8A). Since GTEx raw data is not available publicly, conditional analysis was not performed to infer colocalization.

Transcription factor motif analysis

We used motif breaker R to detect the disruption of the transcription motif binding site where there was a variant within a chromatin accessibility peak (Figure 5D)[111].

TWAS analysis

We performed transcriptome wide association analysis for progenitor and neurons separately with FUSION software (http://gusevlab.org/projects/fusion/,[61]). First, we obtained a set of variants shared between the genotypes from 1000 Genome European phase 3[82] and our study population restricted to variants described for eQTL analysis, and removed monomorphic variants within European genotype data. We estimated cis-heritability of genes (including variants within 1 MB +/- window of the TSS) and intron junctions (including variants within 200kb +/- window of two ends of intron junctions) with GCTA software[112] by controlling for same covariates for global gene expression/splicing and 10 PCs of global genotypes used in e/sQTL analysis. VST normalized gene expressions were further subject to quantile normalization for heritability estimation. 1,703/973 genes and 6,728/6,799 intron junctions were significantly cis-heritable in progenitors/neurons for heritability p-value < 0.01. To determine the method to be used to estimate genetic component of gene expression/splicing (weights), we performed leave-one-out cross validation[113] for the prediction models including LASSO regression[114], Elastic-net regression[115] and EMMAX[89] within FUSION software. We used the weights computed from the prediction model with the highest cross validation R^2 (the highest performance) per gene/intron junction for downstream analysis for progenitor, neuron and fetal bulk brain tissue. For adult brain bulk tissue data, we obtained the weights of genes and intron junctions from CommonMind Consortium study[63].

Before running TWAS analysis, we prepared GWAS summary statistics for schizophrenia (SCZ)[1], major depression disorder (MDD)[101], educational attainment (EA)[102], neuroticism[103], IQ[116], Alzheimer's disease (AD)[117], Parkinson's disease (PD)[106], and Global Surface Area (GSA) and average thickness from ENIGMA study[5] with following adaptations: (1) we obtained common variants found both in genotype files from our study and in GWAS summary statistics; (2) we calculated z-score by dividing the beta coefficient by the standard error if the beta coefficient was available in the summary statistics, or dividing the natural logarithm of odds ratio by the standard error if odds ratio was given in the summary statistics; (3) The sign of the z-score were matched based on the allelic directionality of weights from FUSION software.

To perform TWAS analysis, we tested the association between the predicted gene expression/splicing (w) and brain traits listed above (Z) by implementing the algorithm $Z_{TWAS} = w'Z/\sqrt{w'Dw}$ where D is the LD matrix as the covariance among all cis-variants from the FUSION software [61,63]. Since the population structure of our dataset was different from European neuropsychiatric GWAS, we performed TWAS analysis separately with different LD estimates computed based on our study or European population from 1000 Genomes Phase 3 as the covariance. For variants missing in GWAS summary statistics which existed in our study's genotypes, we implemented IMPG imputation[118] allowing 40% of missing variants as maximum ratio within the FUSION algorithm.

To identify genes/intron junctions not driven by co-expression, we defined jointly independent genes/intron junctions through performing summary-statistic-based joint analysis[62], where we replaced SNPs with genes/intron junctions as described in previous work[63] within the FUSION software. Implementing genes/intron junctions to the model one at a time in decreasing order of significance, we evaluated whether the conditional TWAS test remained significant. Those with significant conditional TWAS association were defined as jointly independent.

References

1. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
2. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
3. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
4. Matoba, N. *et al.* Common genetic risk variants identified in the SPARK cohort implicate DDHD2 as a novel autism risk gene. *medRxiv* (2020).
5. Grasby, K. L. *et al.* The genetic architecture of the human cerebral cortex. *Science* **367**, (2020).
6. Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018).
7. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
8. Fraser, H. B. & Xie, X. Common polymorphic transcript variation in human disease. *Genome Res.* **19**, 567–575 (2009).
9. Li, Y. I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).
10. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, (2018).
11. Takata, A., Matsumoto, N. & Kato, T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* **8**, 14519 (2017).
12. Xu, B. *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in

schizophrenia. *Nat. Genet.* **44**, 1365–1369 (2012).

13. Raj, B. & Blencowe, B. J. Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron* **87**, 14–27 (2015).
14. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
15. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, (2018).
16. Liang, D. *et al.* Cell-type specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Biorxiv* (2020) doi:10.1101/2020.01.13.904862.
17. de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289–304.e18 (2018).
18. O'Brien, H. E. *et al.* Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* **19**, 194 (2018).
19. Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* vol. 181 745 (2020).
20. Werling, D. M. *et al.* Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic Coordination in the Developing Human Prefrontal Cortex. *Cell Rep.* **31**, 107489 (2020).
21. Cuomo, A. S. E. *et al.* Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020).
22. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
23. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
24. Liang, D., Elwell, A. L., Aygün, N., Lafferty, M. J. & Krupa, O. Cell-type specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *bioRxiv*

(2020).

25. Rosen, E. Y. *et al.* Functional genomic analyses identify pathways dysregulated by progranulin deficiency, implicating Wnt signaling. *Neuron* **71**, 1030–1042 (2011).
26. Stein, J. L. *et al.* A quantitative framework to evaluate modeling of cortical development by neural stem cells. *Neuron* **83**, 69–86 (2014).
27. Konopka, G. *et al.* Modeling the functional genomics of autism using human neurons. *Mol. Psychiatry* **17**, 202–214 (2012).
28. Palmer, T. D. *et al.* Progenitor cells from human brain after death. *Nature* **411**, 42–43 (2001).
29. Hansen, D. V., Lui, J. H., Parker, P. R. L. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
30. Gómez-López, S. *et al.* Sox2 and Pax6 maintain the proliferative and developmental potential of gliogenic neural stem cells In vitro. *Glia* **59**, 1588–1599 (2011).
31. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785–801.e8 (2019).
32. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
33. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
34. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).
35. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
36. Davis, J. R. *et al.* An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am. J. Hum. Genet.* **98**, 216–224 (2016).
37. Huang, Q. Q., Ritchie, S. C., Brozynska, M. & Inouye, M. Power, false discovery rate and

Winner's Curse in eQTL studies. *Nucleic Acids Research* vol. 46 e133–e133 (2018).

38. Strunz, T. *et al.* A mega-analysis of expression quantitative trait loci (eQTL) provides insight into the regulatory architecture of gene expression variation in liver. *Sci. Rep.* **8**, 5865 (2018).
39. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
40. Stranger, B. E. *et al.* Patterns of Cis Regulatory Variation in Diverse Human Populations. *PLoS Genetics* vol. 8 e1002639 (2012).
41. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
42. Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* **16**, 195 (2015).
43. Pastinen, T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet.* **11**, 533–538 (2010).
44. Gonzalez, N., Moody, T. W., Igarashi, H., Ito, T. & Jensen, R. T. Bombesin-related peptides and their receptors: recent advances in their role in physiology and disease states. *Curr. Opin. Endocrinol. Diabetes Obes.* **15**, 58–64 (2008).
45. Reay, W. R. & Cairns, M. J. Pairwise common variant meta-analyses of schizophrenia with other psychiatric disorders reveals shared and distinct gene and gene-set associations. *Transl. Psychiatry* **10**, 134 (2020).
46. Hall, L. S. *et al.* Cis-effects on gene expression in the human prenatal brain associated with genetic risk for neuropsychiatric disorders. *Mol. Psychiatry* (2020) doi:10.1038/s41380-020-0743-3.
47. Monlong, J., Calvo, M., Ferreira, P. G. & Guigó, R. Identification of genetic variants associated with alternative splicing using sQTLseekeR. *Nat. Commun.* **5**, 4698 (2014).

48. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018).

49. Zheng, S. *et al.* PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. *Nat. Neurosci.* **15**, 381–8, S1 (2012).

50. Zheng, S. Alternative splicing and nonsense-mediated mRNA decay enforce neural specific gene expression. *Int. J. Dev. Neurosci.* **55**, 102–108 (2016).

51. Guo, J. *et al.* Developmental disruptions underlying brain abnormalities in ciliopathies. *Nat. Commun.* **6**, 7857 (2015).

52. Yang, Y.-C. T. *et al.* CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics* **16**, 51 (2015).

53. Herrlinger, S. *et al.* Lin28-mediated temporal promotion of protein synthesis is crucial for neural progenitor cell maintenance and brain development in mice. *Development* **146**, (2019).

54. Pfister, J. A. & D'Mello, S. R. Regulation of Neuronal Survival by Nucleophosmin 1 (NPM1) Is Dependent on Its Expression Level, Subcellular Localization, and Oligomerization Status. *J. Biol. Chem.* **291**, 20787–20797 (2016).

55. Civelek, M. *et al.* Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am. J. Hum. Genet.* **100**, 428–443 (2017).

56. McKinley, K. L. & Cheeseman, I. M. Large-Scale Analysis of CRISPR/Cas9 Cell-Cycle Knockouts Reveals the Diversity of p53-Dependent Responses to Cell-Cycle Defects. *Dev. Cell* **40**, 405–420.e2 (2017).

57. Lee, S., Koh, W., Kim, H.-T., Kim, C.-H. & Lee, S. Cancer-upregulated gene 2 (CUG2) overexpression induces apoptosis in SKOV-3 cells. *Cell Biochemistry and Function* vol. 28 461–468 (2010).

58. Li, M. *et al.* A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. *Nat. Med.* **22**, 649–656 (2016).

59. Peter, C. J. *et al.* In vivo epigenetic editing of Sema6a promoter reverses transcallosal dysconnectivity caused by C11orf46/Arl14ep risk gene. *Nat. Commun.* **10**, 4112 (2019).

60. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics* vol. 47 1091–1098 (2015).

61. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

62. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–75, S1–3 (2012).

63. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nature Genetics* vol. 50 538–548 (2018).

64. Hedberg, C., Oldfors, A. & Darin, N. B3GALNT2 is a gene associated with congenital muscular dystrophy with brain malformations. *Eur. J. Hum. Genet.* **22**, 707–710 (2014).

65. Maroofian, R. *et al.* B3GALNT2 mutations associated with non-syndromic autosomal recessive intellectual disability reveal a lack of genotype-phenotype associations in the muscular dystrophy-dystroglycanopathies. *Genome Med.* **9**, 118 (2017).

66. Garone, C. *et al.* Defective mitochondrial rRNA methyltransferase MRM2 causes MELAS-like clinical syndrome. *Hum. Mol. Genet.* **26**, 4257–4266 (2017).

67. Freude, K. *et al.* Mutations in the FTSJ1 gene coding for a novel S-adenosylmethionine-binding protein cause nonsyndromic X-linked mental retardation. *Am. J. Hum. Genet.* **75**, 305–309 (2004).

68. Kim, H. *et al.* Cancer-upregulated gene 2 (CUG2), a new component of centromere complex, is required for kinetochore function. *Mol. Cells* **27**, 697–701 (2009).

69. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).

70. Park, Y., Sarkar, A., Nguyen, K. & Kellis, M. Causal Mediation Analysis Leveraging Multiple

Types of Summary Statistics Data. *arXiv [stat.ML]* (2019).

71. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
72. Harrison, P. J. Using our brains: the findings, flaws, and future of postmortem studies of psychiatric disorders. *Biol. Psychiatry* **69**, 102–103 (2011).
73. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
74. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
75. Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nature Genetics* vol. 51 343–353 (2019).
76. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
77. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. 29, 15--21. (2013).
78. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
79. van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063 (2015).
80. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
81. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
82. Consortium, T. 1000 G. P. & The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* vol. 526 68–74 (2015).

83. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

84. Jun, G. *et al.* Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848 (2012).

85. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

86. Reimand, J. *et al.* g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research* vol. 44 W83–W89 (2016).

87. Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38**, e169–e169 (2010).

88. Cahill, K. M., Huo, Z., Tseng, G. C., Logan, R. W. & Seney, M. L. Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. *Sci. Rep.* **8**, 9588 (2018).

89. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).

90. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).

91. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).

92. Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am. J. Hum. Genet.* **102**, 1169–1184 (2018).

93. Jansen, R. *et al.* Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.* **26**, 1444–1451 (2017).

94. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).

95. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* **33**, 364–376 (2015).

96. Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Tools and best practices for allelic expression analysis. bioRχiv. (2015).

97. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

98. Feng, Y.-Y. *et al.* RegTools: Integrated analysis of genomic and transcriptomic data for discovery of splicing variants in cancer. doi:10.1101/436634.

99. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* **8**, e1002555 (2012).

100. Rosenblatt, J. D. & Stein, J. L. RRHO: test overlap using the rank-rank hypergeometric test. R package version 1.22. 0. (2014).

101. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

102. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).

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

104. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for ADHD. bioRxiv (2017). *DOI* **10**, 145581.

105. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).

106. Nalls, M. A. *et al.* Parkinson's disease genetics: identifying novel risk loci, providing causal

insights and improving estimates of heritable risk. *BioRxiv* 388165 (2018).

107. Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).

108. International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat. Commun.* **9**, 5269 (2018).

109. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).

110. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

111. Coetzee, S. G., Coetzee, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847–3849 (2015).

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

113. Meijer, R. J. & Goeman, J. J. Efficient approximate k-fold and leave-one-out cross-validation for ridge regression. *Biom. J.* **55**, 141–155 (2013).

114. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **58**, 267–288 (1996).

115. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.* **67**, 301–320 (2005).

116. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).

117. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).

118. Pasaniuc, B. *et al.* Fast and accurate imputation of summary statistics enhances evidence

of functional enrichment. *Bioinformatics* **30**, 2906–2914 (2014).

Figure 1

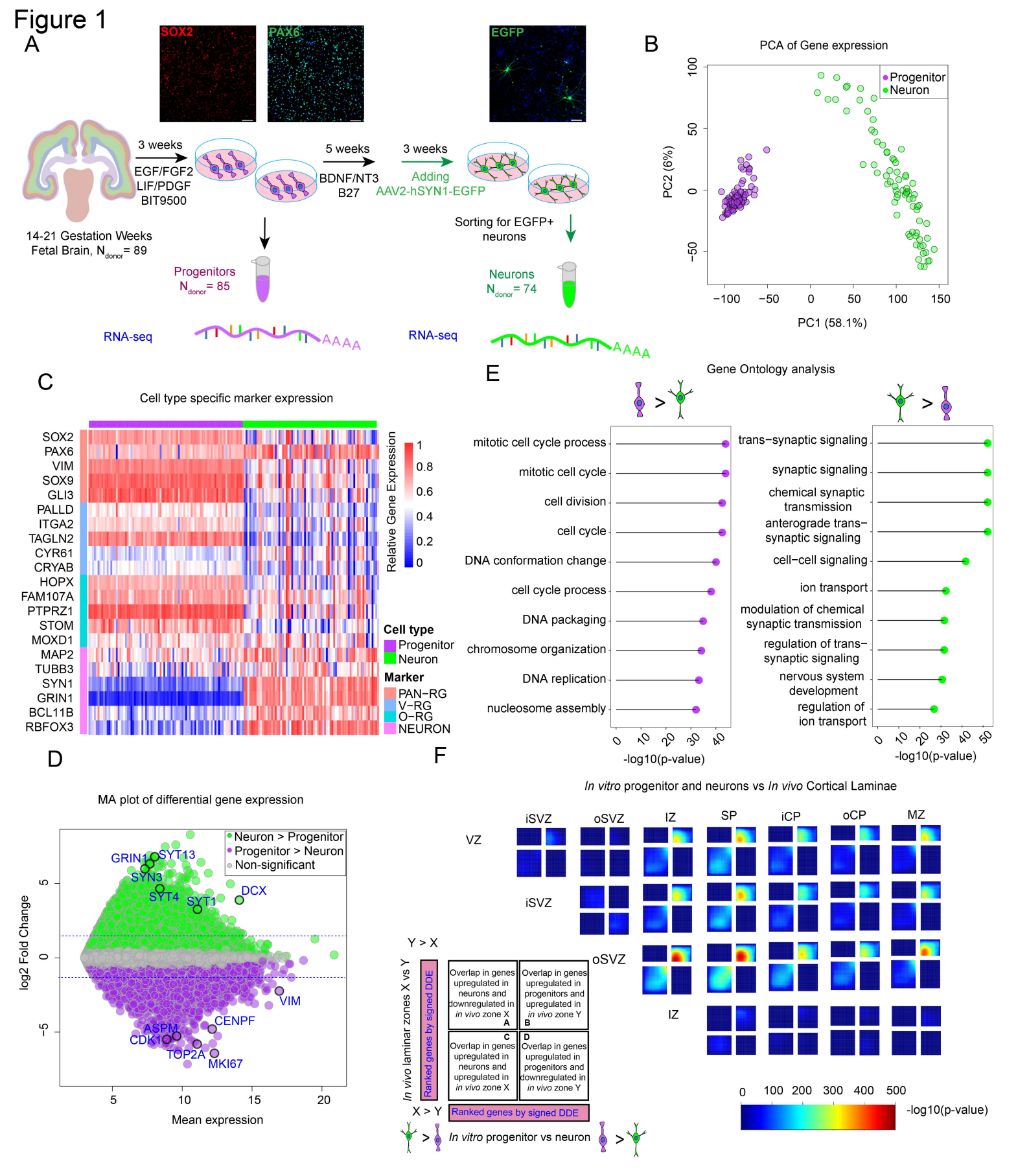

A
SOX2
PAX6
EGFP
3 weeks
EGF/FGF2
LIF/PDGF
BIT9500
5 weeks
BDNF/NT3
B27
3 weeks
Adding
AAV2-hSYN1-EGFP
14-21 Gestation Weeks
Fetal Brain, N_donor = 89
Progenitors
N_donor = 85
Sorting for EGFP+
neurons
Neurons
N_donor = 74
RNA-seq
RNA-seq
B
PCA of Gene expression
Progenitor
Neuron
PC2 (6%)
PC1 (58.1%)
C
Cell type specific marker expression
SOX2
PAX6
VIM
SOX9
GLI3
PALLD
ITGA2
TAGLN2
CYR61
CRYAB
HOPX
FAM107A
PTPRZ1
STOM
MOXD1
MAP2
TUBB3
SYN1
GRIN1
BCL11B
RBFOX3
Relative Gene Expression
Cell type
Progenitor
Neuron
Marker
PAN−RG
V−RG
O−RG
NEURON
D
MA plot of differential gene expression
Neuron > Progenitor
Progenitor > Neuron
Non-significant
GRIN1
SYT13
SYN3
SYT4
SYT1
DCX
VIM
CENPF
ASPM
CDK1
TOP2A
MKI67
log2 Fold Change
Mean expression
E
Gene Ontology analysis
mitotic cell cycle process
mitotic cell cycle
cell division
cell cycle
DNA conformation change
cell cycle process
DNA packaging
chromosome organization
DNA replication
nucleosome assembly
-log10(p-value)
trans−synaptic signaling
synaptic signaling
chemical synaptic transmission
anterograde trans−synaptic signaling
cell−cell signaling
ion transport
modulation of chemical synaptic transmission
regulation of trans−synaptic signaling
nervous system development
regulation of ion transport
-log10(p-value)
F
In vitro progenitor and neurons vs In vivo Cortical Laminae
iSVZ
oSVZ
IZ
SP
iCP
oCP
MZ
VZ
iSVZ
oSVZ
IZ
Y > X
X > Y
In vivo laminar zones X vs Y
Ranked genes by signed DDE
Overlap in genes upregulated in neurons and downregulated in in vivo zone X
A
Overlap in genes upregulated in progenitors and upregulated in in vivo zone Y
B
Overlap in genes upregulated in neurons and upregulated in in vivo zone X
C
Overlap in genes upregulated in progenitors and downregulated in in vivo zone Y
D
Ranked genes by signed DDE
In vitro progenitor vs neuron
0
100
200
300
400
500
-log10(p-value)

Figure 2

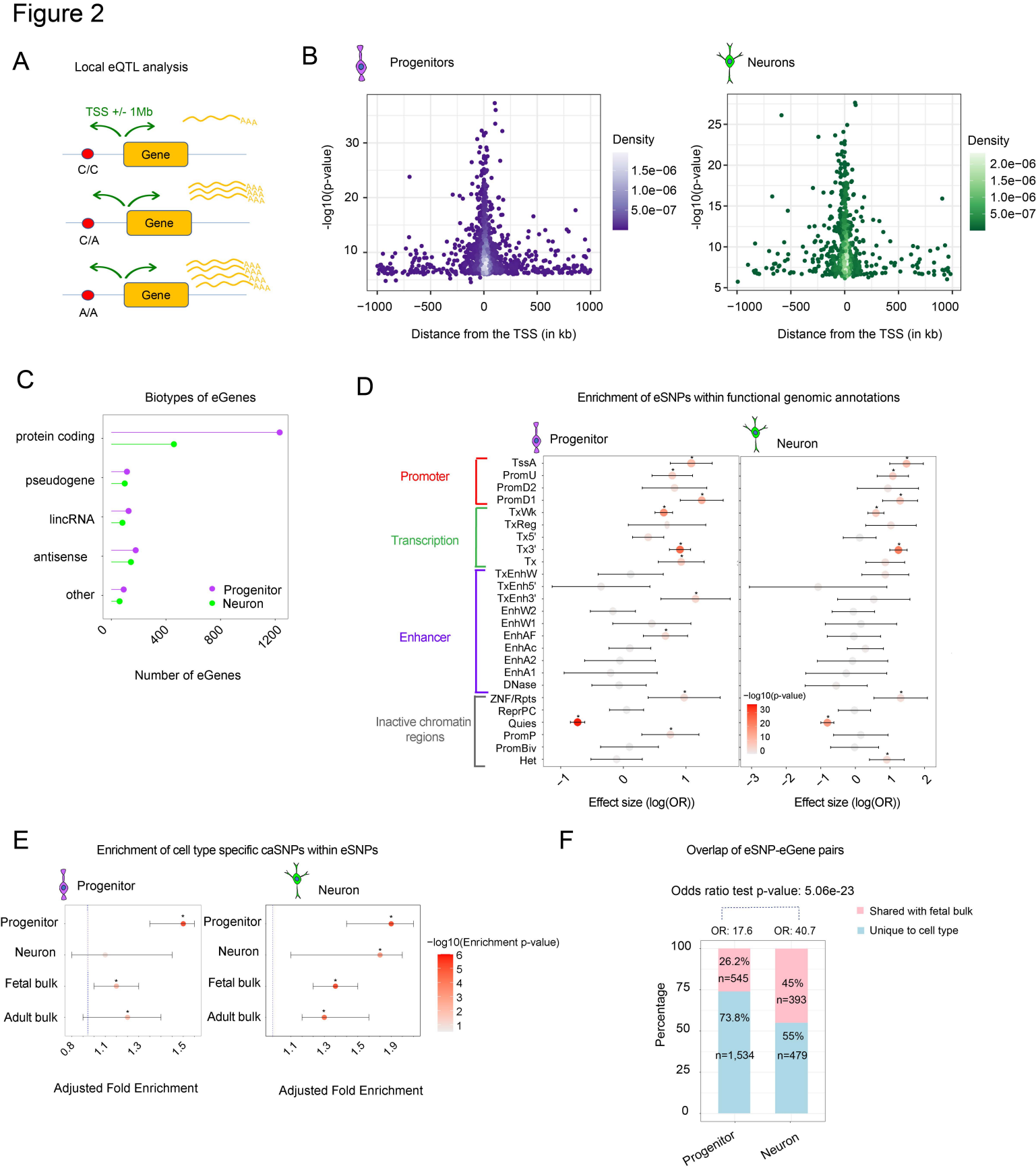

A
Local eQTL analysis
TSS +/- 1Mb
Gene
C/C
C/A
A/A
B
Progenitors
Neurons
-log10(p-value)
Density
Distance from the TSS (in kb)
C
Biotypes of eGenes
protein coding
pseudogene
lincRNA
antisense
other
Progenitor
Neuron
Number of eGenes
D
Enrichment of eSNPs within functional genomic annotations
Progenitor
Neuron
Promoter
Transcription
Enhancer
Inactive chromatin regions
TssA
PromU
PromD2
PromD1
TxWk
TxReg
Tx5'
Tx3'
Tx
TxEnhW
TxEnh5'
TxEnh3'
EnhW2
EnhW1
EnhAF
EnhAc
EnhA2
EnhA1
DNase
ZNF/Rpts
ReprPC
Quies
PromP
PromBiv
Het
-log10(p-value)
Effect size (log(OR))
E
Enrichment of cell type specific caSNPs within eSNPs
Progenitor
Neuron
Fetal bulk
Adult bulk
-log10(Enrichment p-value)
Adjusted Fold Enrichment
F
Overlap of eSNP-eGene pairs
Odds ratio test p-value: 5.06e-23
OR: 17.6
OR: 40.7
Shared with fetal bulk
Unique to cell type
Percentage
26.2%
n=545
73.8%
n=1,534
45%
n=393
55%
n=479
Progenitor
Neuron

Figure 3

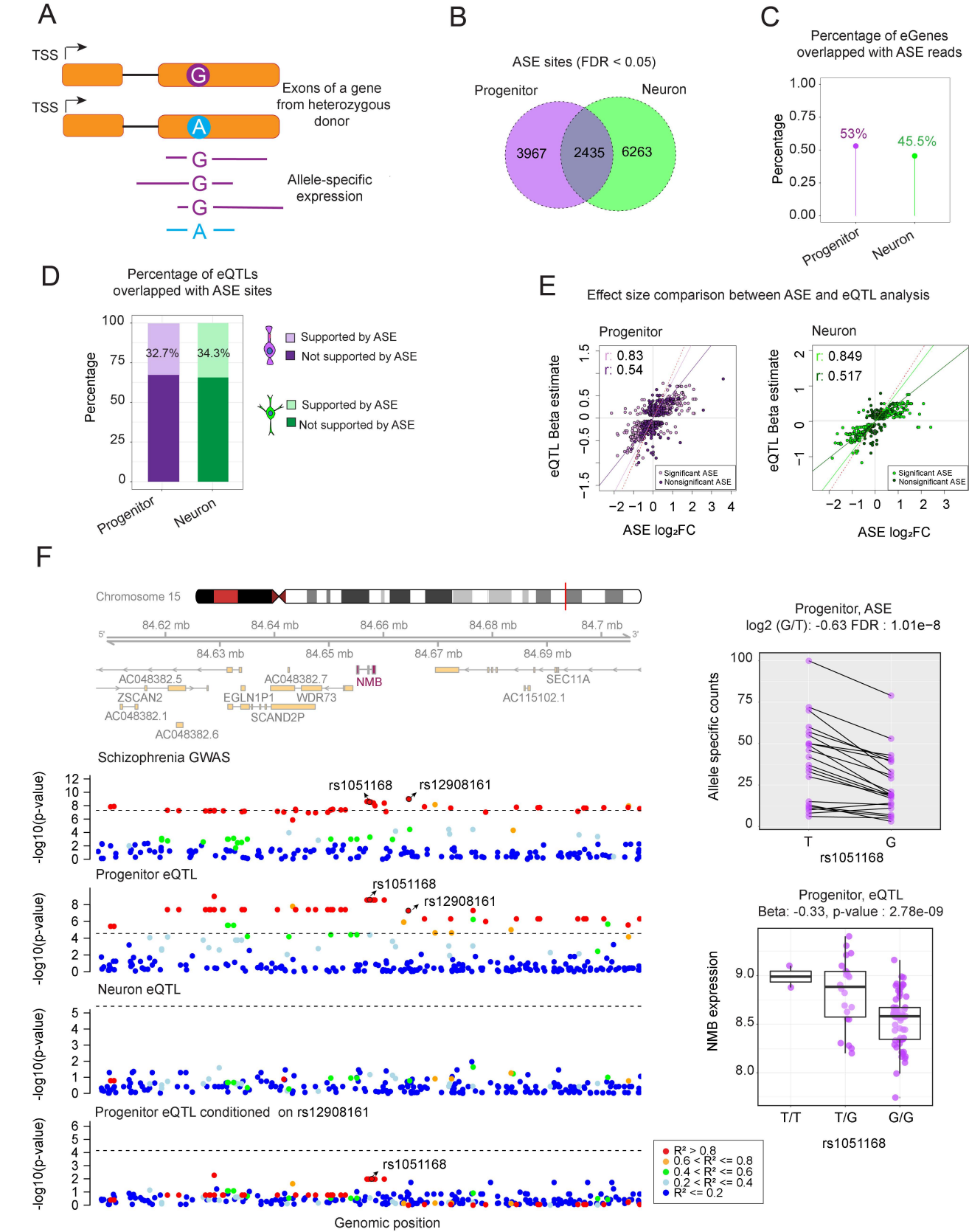

A
TSS
TSS
G
A
Exons of a gene from heterozygous donor
Allele-specific expression
B
ASE sites (FDR < 0.05)
Progenitor
Neuron
3967
2435
6263
C
Percentage of eGenes overlapped with ASE reads
Percentage
53%
45.5%
Progenitor
Neuron
D
Percentage of eQTLs overlapped with ASE sites
Percentage
32.7%
34.3%
Supported by ASE
Not supported by ASE
Supported by ASE
Not supported by ASE
Progenitor
Neuron
E
Effect size comparison between ASE and eQTL analysis
Progenitor
r: 0.83
r: 0.54
Significant ASE
Nonsignificant ASE
eQTL Beta estimate
ASE log2FC
Neuron
r: 0.849
r: 0.517
Significant ASE
Nonsignificant ASE
eQTL Beta estimate
ASE log2FC
F
Chromosome 15
84.62 mb
84.64 mb
84.66 mb
84.68 mb
84.7 mb
84.63 mb
84.65 mb
84.67 mb
84.69 mb
AC048382.5
ZSCAN2
AC048382.1
AC048382.6
AC048382.7
EGLN1P1
WDR73
SCAND2P
NMB
SEC11A
AC115102.1
Schizophrenia GWAS
rs1051168
rs12908161
-log10(p-value)
Progenitor eQTL
rs1051168
rs12908161
-log10(p-value)
Neuron eQTL
-log10(p-value)
Progenitor eQTL conditioned on rs12908161
rs1051168
-log10(p-value)
Genomic position
R² > 0.8
0.6 < R² <= 0.8
0.4 < R² <= 0.6
0.2 < R² <= 0.4
R² <= 0.2
Progenitor, ASE
log2 (G/T): -0.63 FDR : 1.01e−8
Allele specific counts
T
G
rs1051168
Progenitor, eQTL
Beta: -0.33, p-value : 2.78e-09
NMB expression
T/T
T/G
G/G
rs1051168

Figure 4

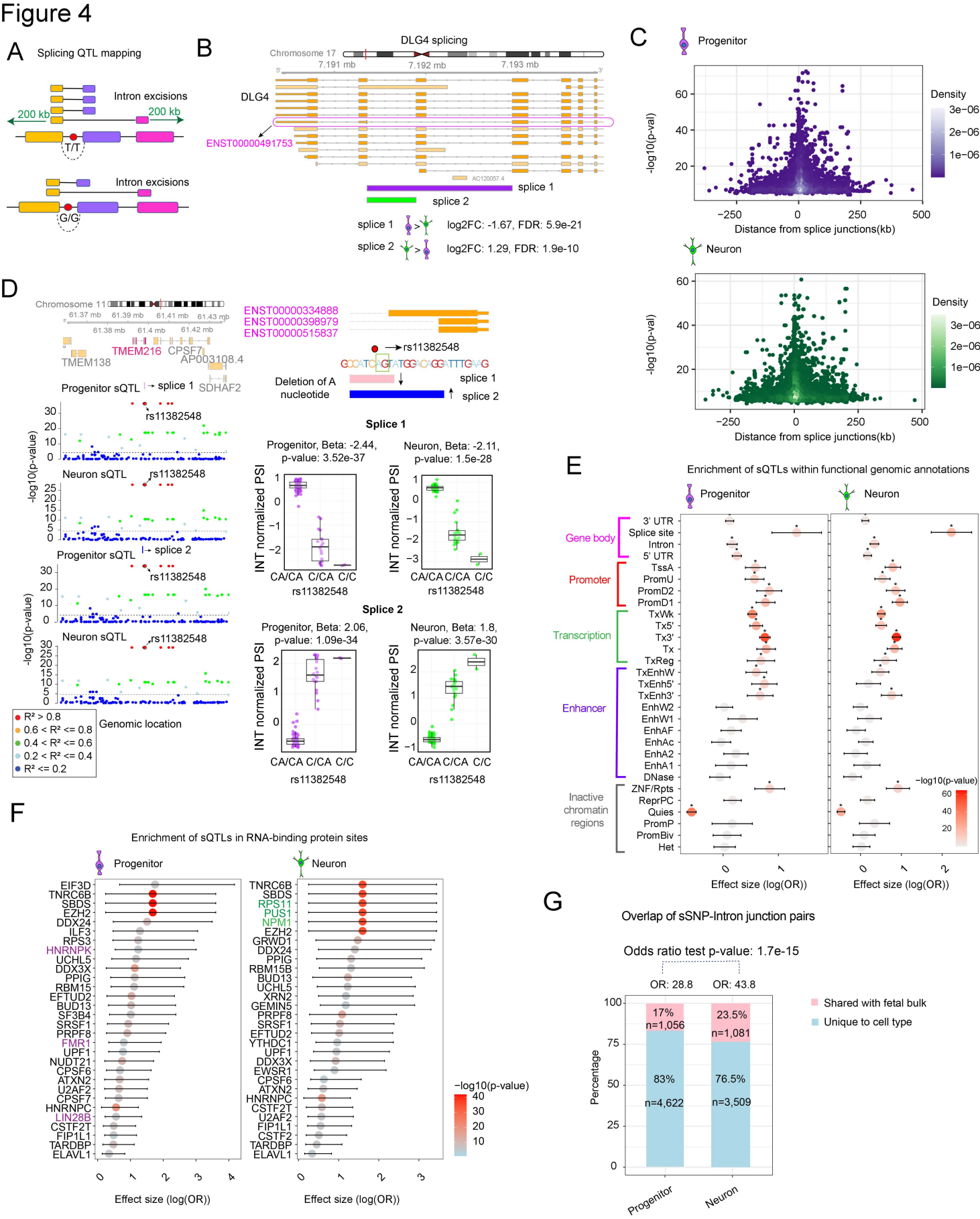

A
Splicing QTL mapping
Intron excisions
200 kb
200 kb
T/T
Intron excisions
G/G
B
DLG4 splicing
Chromosome 17
7.191 mb
7.192 mb
7.193 mb
DLG4
ENST00000491753
splice 1
splice 2
splice 1
log2FC: -1.67, FDR: 5.9e-21
splice 2
log2FC: 1.29, FDR: 1.9e-10
C
Progenitor
Density
3e-06
2e-06
1e-06
-log10(p-val)
Distance from splice junctions(kb)
Neuron
Density
3e-06
2e-06
1e-06
-log10(p-val)
Distance from splice junctions(kb)
D
Chromosome 11
TMEM138
TMEM216
CPSF7
AP003108.4
SDHAF2
ENST00000334888
ENST00000398979
ENST00000515837
rs11382548
Deletion of A nucleotide
splice 1
splice 2
Progenitor sQTL
splice 1
rs11382548
Neuron sQTL
rs11382548
Progenitor sQTL
splice 2
rs11382548
Neuron sQTL
rs11382548
-log10(p-value)
Genomic location
R² > 0.8
0.6 < R² <= 0.8
0.4 < R² <= 0.6
0.2 < R² <= 0.4
R² <= 0.2
Splice 1
Progenitor, Beta: -2.44, p-value: 3.52e-37
Neuron, Beta: -2.11, p-value: 1.5e-28
INT normalized PSI
CA/CA C/CA C/C
rs11382548
Splice 2
Progenitor, Beta: 2.06, p-value: 1.09e-34
Neuron, Beta: 1.8, p-value: 3.57e-30
INT normalized PSI
CA/CA C/CA C/C
rs11382548
E
Enrichment of sQTLs within functional genomic annotations
Progenitor
Neuron
Gene body
3' UTR
Splice site
Intron
5' UTR
Promoter
TssA
PromU
PromD2
PromD1
Transcription
TxWk
Tx5'
Tx3'
Tx
TxReg
Enhancer
TxEnhW
TxEnh5'
TxEnh3'
EnhW2
EnhW1
EnhAF
EnhAc
EnhA2
EnhA1
DNase
Inactive chromatin regions
ZNF/Rpts
ReprPC
Quies
PromP
PromBiv
Het
-log10(p-value)
Effect size (log(OR))
Effect size (log(OR))
F
Enrichment of sQTLs in RNA-binding protein sites
Progenitor
Neuron
EIF3D
TNRC6B
SBDS
EZH2
DDX24
ILF3
RPS3
HNRNPK
UCHL5
DDX3X
PPIG
RBM15
EFTUD2
BUD13
SF3B4
SRSF1
PRPF8
FMR1
UPF1
NUDT21
CPSF6
ATXN2
U2AF2
CPSF7
HNRNPC
LIN28B
CSTF2T
FIP1L1
TARDBP
ELAVL1
TNRC6B
SBDS
RPS11
PUS1
NPM1
EZH2
GRWD1
DDX24
PPIG
RBM15B
BUD13
UCHL5
XRN2
GEMIN5
PRPF8
SRSF1
EFTUD2
YTHDC1
UPF1
DDX3X
EWSR1
CPSF6
ATXN2
HNRNPC
CSTF2T
U2AF2
FIP1L1
CSTF2
TARDBP
ELAVL1
-log10(p-value)
Effect size (log(OR))
Effect size (log(OR))
G
Overlap of sSNP-Intron junction pairs
Odds ratio test p-value: 1.7e-15
OR: 28.8
OR: 43.8
17%
n=1,056
23.5%
n=1,081
83%
n=4,622
76.5%
n=3,509
Percentage
Shared with fetal bulk
Unique to cell type
Progenitor
Neuron

Figure 5

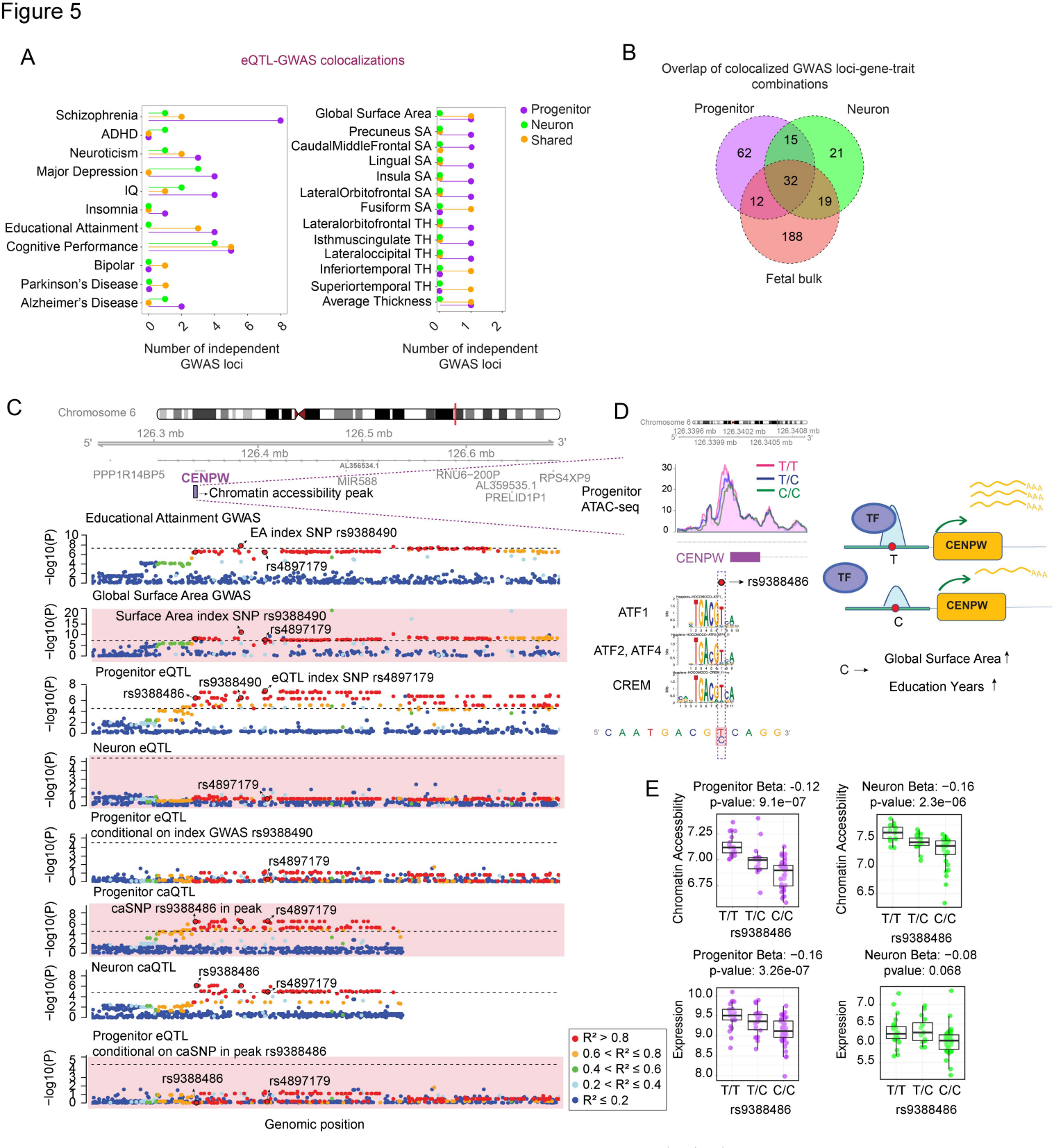

A
eQTL-GWAS colocalizations
Schizophrenia
ADHD
Neuroticism
Major Depression
IQ
Insomnia
Educational Attainment
Cognitive Performance
Bipolar
Parkinson's Disease
Alzheimer's Disease
Number of independent GWAS loci
Global Surface Area
Precuneus SA
CaudalMiddleFrontal SA
Lingual SA
Insula SA
LateralOrbitofrontal SA
Fusiform SA
Lateralorbitofrontal TH
Isthmuscingulate TH
Lateraloccipital TH
Inferiortemporal TH
Superiortemporal TH
Average Thickness
Number of independent GWAS loci
Progenitor
Neuron
Shared
B
Overlap of colocalized GWAS loci-gene-trait combinations
Progenitor
Neuron
Fetal bulk
62
15
21
32
12
19
188
C
Chromosome 6
126.3 mb
126.5 mb
126.4 mb
126.6 mb
PPP1R14BP5
CENPW
MIR588
RNU6-200P
AL359535.1
PRELID1P1
RPS4XP9
Chromatin accessibility peak
Progenitor ATAC-seq
Educational Attainment GWAS
EA index SNP rs9388490
rs4897179
Global Surface Area GWAS
Surface Area index SNP rs9388490
rs4897179
Progenitor eQTL
rs9388486
rs9388490
eQTL index SNP rs4897179
Neuron eQTL
rs4897179
Progenitor eQTL conditional on index GWAS rs9388490
rs4897179
Progenitor caQTL
caSNP rs9388486 in peak
rs4897179
Neuron caQTL
rs9388486
rs4897179
Progenitor eQTL conditional on caSNP in peak rs9388486
rs9388486
rs4897179
−log10(P)
Genomic position
R² > 0.8
0.6 < R² ≤ 0.8
0.4 < R² ≤ 0.6
0.2 < R² ≤ 0.4
R² ≤ 0.2
D
T/T
T/C
C/C
CENPW
rs9388486
ATF1
ATF2, ATF4
CREM
5' C A A T G A C G T C A G G 3'
TF
CENPW
C →
Global Surface Area ↑
Education Years ↑
E
Progenitor Beta: -0.12
p-value: 9.1e−07
Neuron Beta: −0.16
p-value: 2.3e−06
Chromatin Accessibility
Progenitor Beta: −0.16
p-value: 3.26e-07
Neuron Beta: −0.08
pvalue: 0.068
Expression
T/T T/C C/C
rs9388486

Figure 6

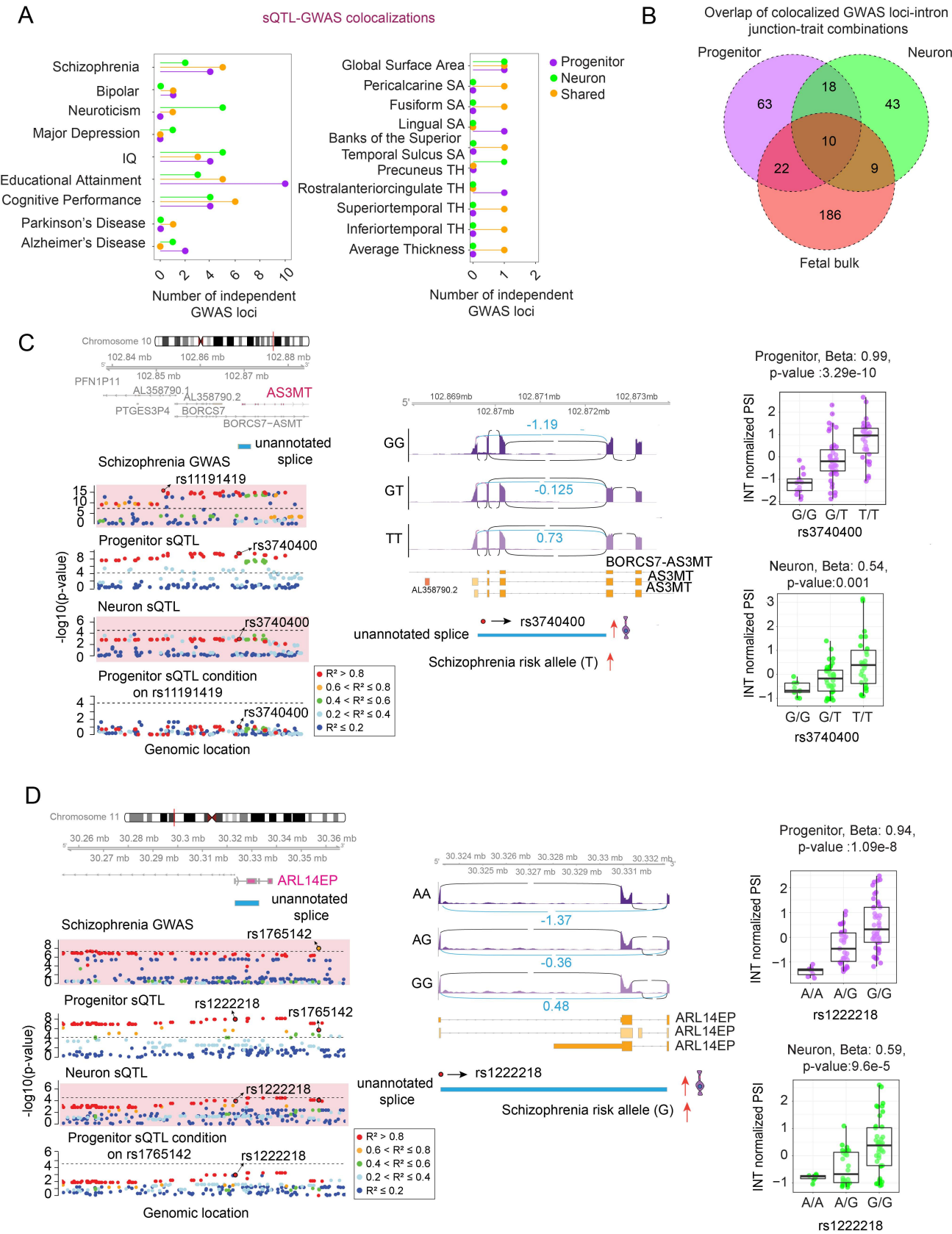
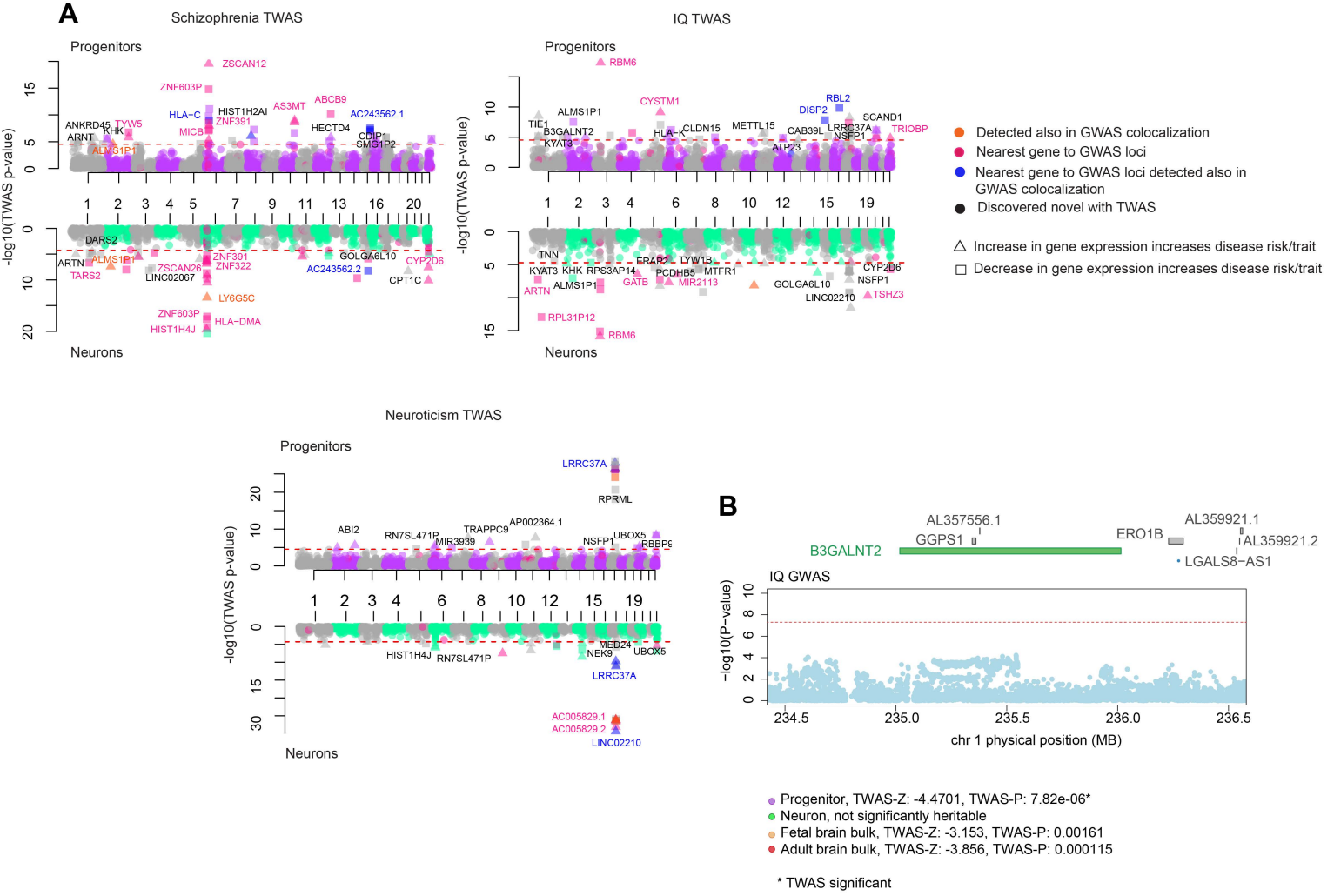

A
sQTL-GWAS colocalizations
Schizophrenia
Bipolar
Neuroticism
Major Depression
IQ
Educational Attainment
Cognitive Performance
Parkinson's Disease
Alzheimer's Disease
Number of independent GWAS loci
Global Surface Area
Pericalcarine SA
Fusiform SA
Lingual SA
Banks of the Superior Temporal Sulcus SA
Precuneus TH
Rostralanteriorcingulate TH
Superiortemporal TH
Inferiortemporal TH
Average Thickness
Progenitor
Neuron
Shared
B
Overlap of colocalized GWAS loci-intron junction-trait combinations
Progenitor
Neuron
Fetal bulk
63
18
43
10
22
9
186
C
Chromosome 10
PFN1P11
AL358790.1
AL358790.2
AS3MT
PTGES3P4
BORCS7
BORCS7-ASMT
unannotated splice
Schizophrenia GWAS
rs11191419
Progenitor sQTL
rs3740400
Neuron sQTL
Progenitor sQTL condition on rs11191419
-log10(p-value)
Genomic location
R² > 0.8
0.6 < R² ≤ 0.8
0.4 < R² ≤ 0.6
0.2 < R² ≤ 0.4
R² ≤ 0.2
GG
GT
TT
-1.19
-0.125
0.73
BORCS7-AS3MT
AS3MT
unannotated splice
rs3740400
Schizophrenia risk allele (T)
Progenitor, Beta: 0.99, p-value :3.29e-10
Neuron, Beta: 0.54, p-value:0.001
INT normalized PSI
G/G G/T T/T
D
Chromosome 11
ARL14EP
unannotated splice
Schizophrenia GWAS
rs1765142
Progenitor sQTL
rs1222218
Neuron sQTL
Progenitor sQTL condition on rs1765142
AA
AG
GG
-1.37
-0.36
0.48
ARL14EP
unannotated splice
rs1222218
Schizophrenia risk allele (G)
Progenitor, Beta: 0.94, p-value :1.09e-8
Neuron, Beta: 0.59, p-value:9.6e-5
A/A A/G G/G

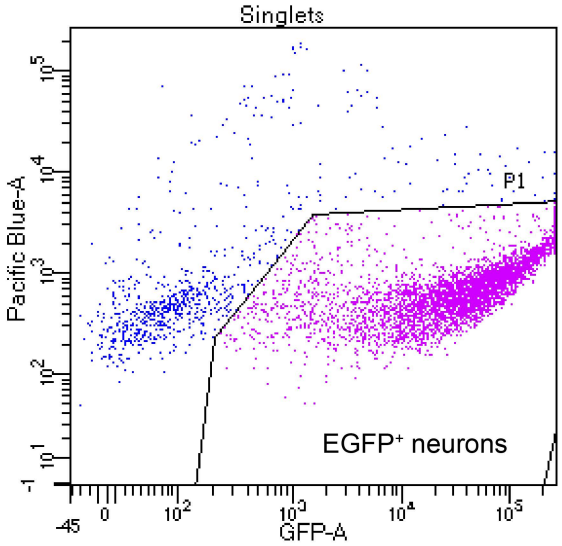
Figure 7


A
Schizophrenia TWAS
Progenitors
Neurons
-log10(TWAS p-value)
ZSCAN12
ZNF603P
HIST1H2AI
HLA-C
MICB
ZNF391
AS3MT
ABCB9
HECTD4
AC243562.1
CDIP1
SMG1P2
ANKRD45
TYW5
KHK
ARNT
ALMS1P1
DARS2
ARTN
TARS2
LINC02067
ZSCAN26
ZNF322
LY6G5C
HIST1H4J
HLA-DMA
GOLGA6L10
AC243562.2
CYP2D6
CPT1C
IQ TWAS
RBM6
CYSTM1
RBL2
DISP2
SCAND1
TIE1
B3GALNT2
KYAT3
HLA-K
CLDN15
METTL15
CAB39L
LRRC37A
NSFP1
TRIOBP
ATP23
TNN
ERAP2
LYW1B
RPS3AP14
PCDHB5
MTFR1
GATB
MIR2113
GOLGA6L10
LINC02210
TSHZ3
RPL31P12
Neuroticism TWAS
LRRC37A
RPRML
ABI2
RN7SL471P
MIR3939
TRAPPC9
AP002364.1
NSFP1
UBOX5
RBBP9
HIST1H4J
RN7SL471P
MED24
NEK9
AC005829.1
AC005829.2
LINC02210
Detected also in GWAS colocalization
Nearest gene to GWAS loci
Nearest gene to GWAS loci detected also in GWAS colocalization
Discovered novel with TWAS
Increase in gene expression increases disease risk/trait
Decrease in gene expression increases disease risk/trait
B
AL357556.1
GGPS1
B3GALNT2
ERO1B
AL359921.1
AL359921.2
LGALS8-AS1
IQ GWAS
−log10(P−value)
chr 1 physical position (MB)
Progenitor, TWAS-Z: -4.4701, TWAS-P: 7.82e-06*
Neuron, not significantly heritable
Fetal brain bulk, TWAS-Z: -3.153, TWAS-P: 0.00161
Adult brain bulk, TWAS-Z: -3.856, TWAS-P: 0.000115
* TWAS significant

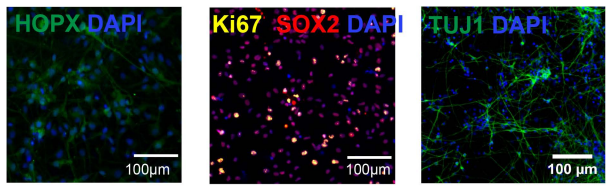
Supplementary Figure 1

A

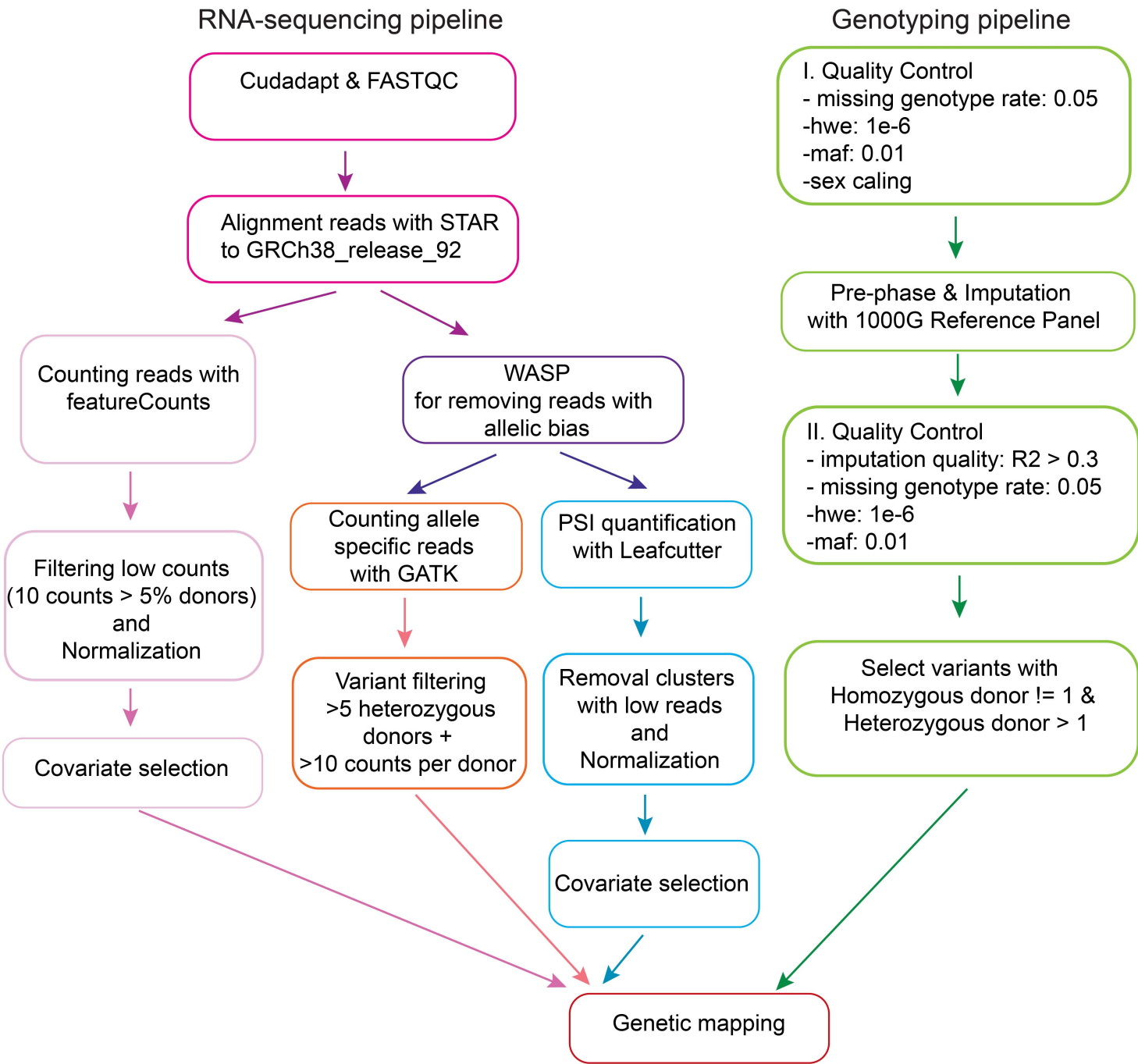
Sorting EGFP labeled neurons



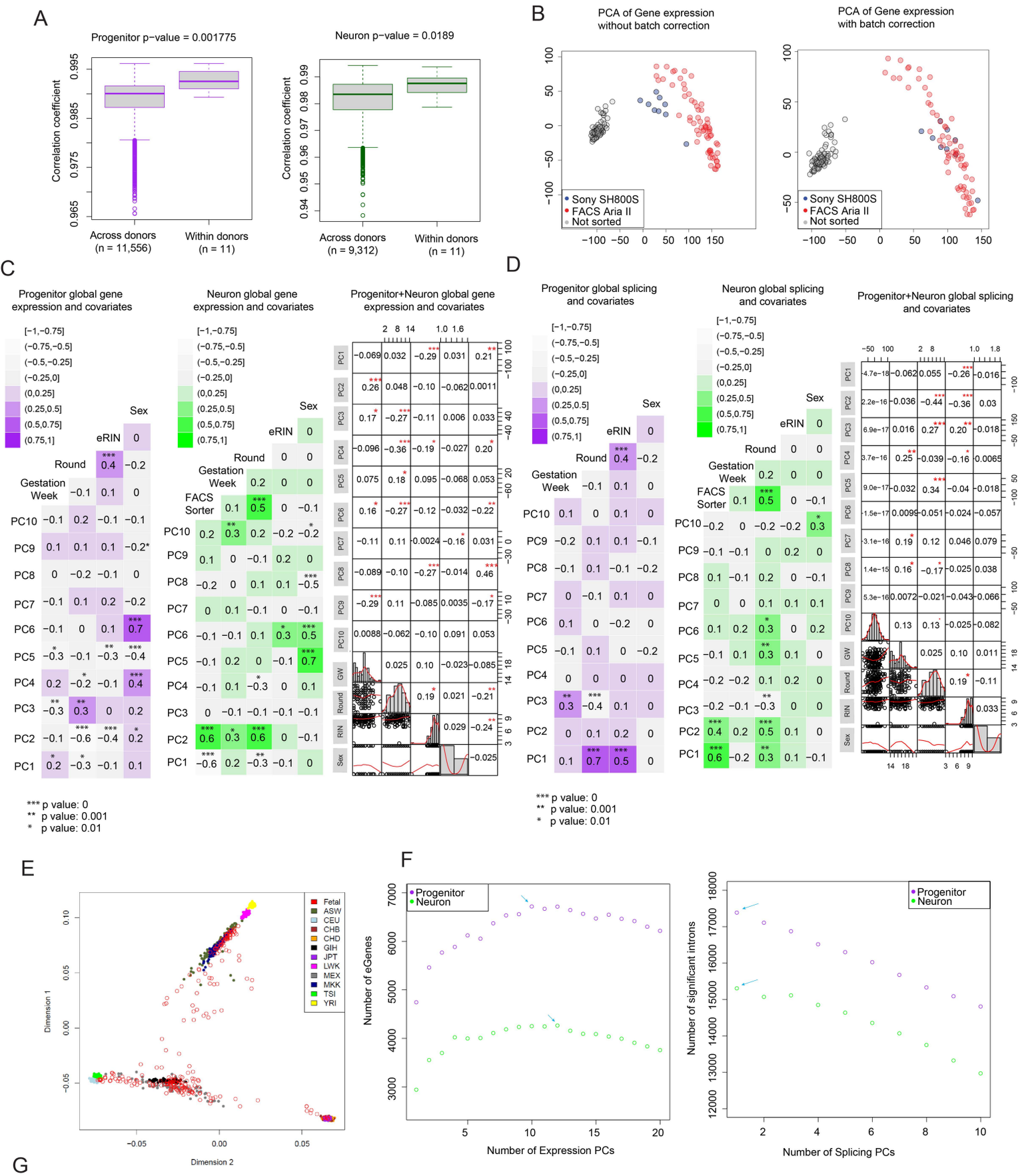
B



C



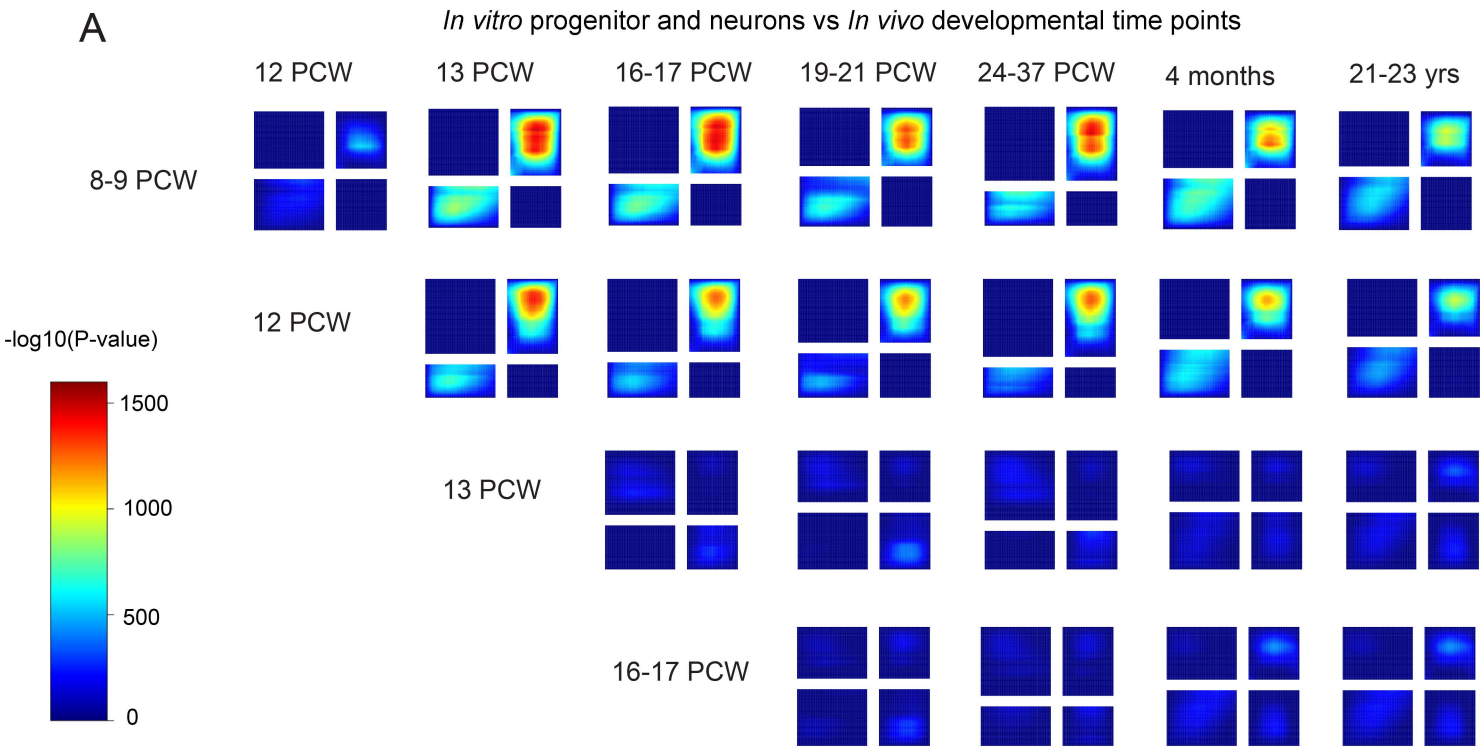
Supplementary Figure 2

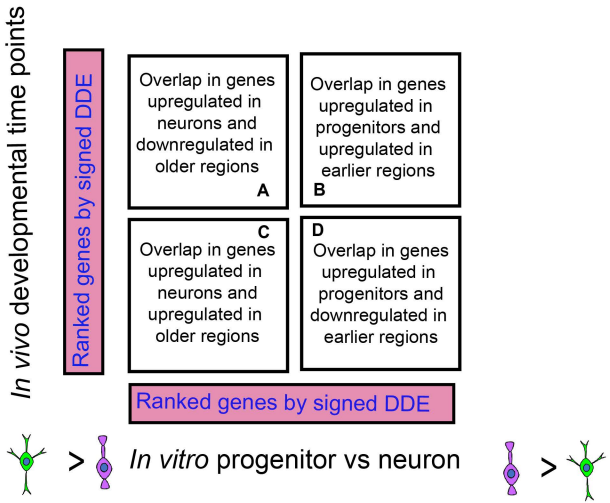


G

	Progenitor λ_{GC}	Neuron λ_{GC}
No control	1.04	1.01
Control for 10 PCs of global expression	1.11	1.05
Control for 10 global genotype PCs +10 PCs of global expression	1.04	1.02
Control for 10 global genotype PCs + 10 PCs of global expression + kinship	1.028	1.007

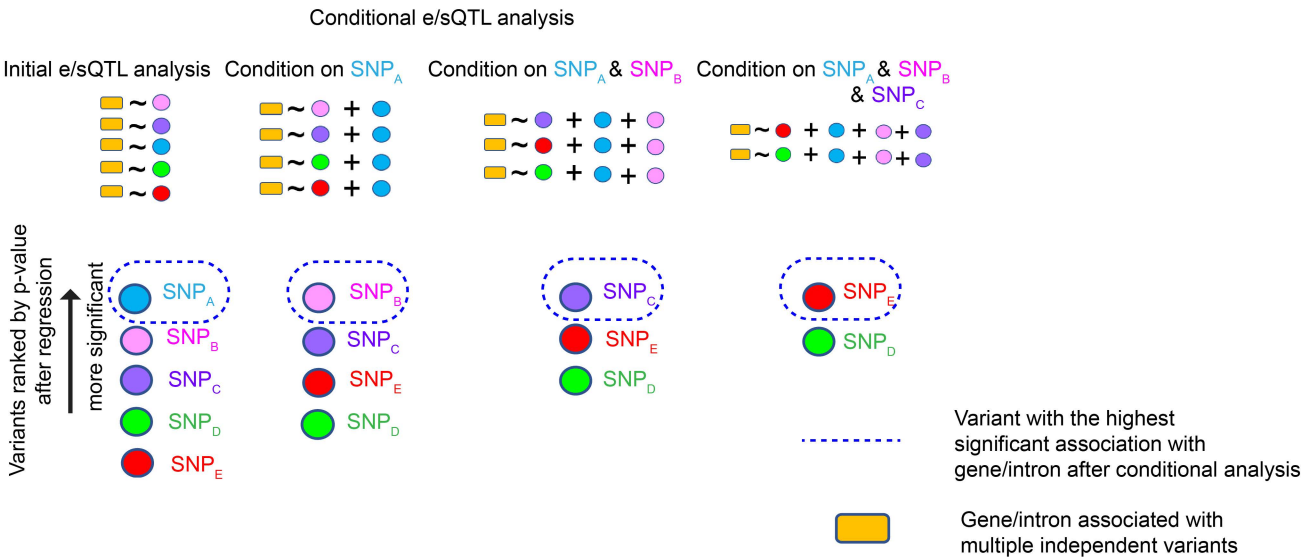
Supplementary Figure 3



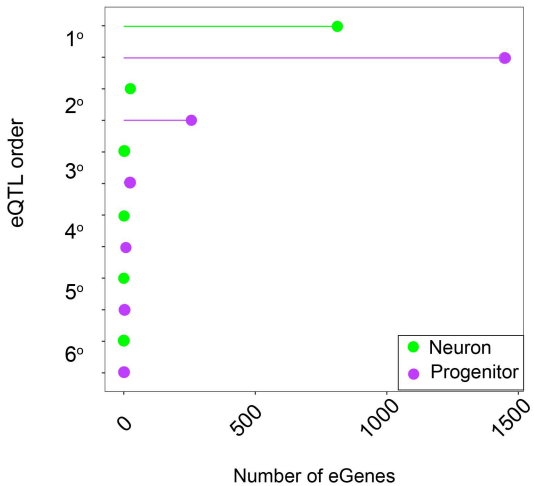


Supplementary Figure 4

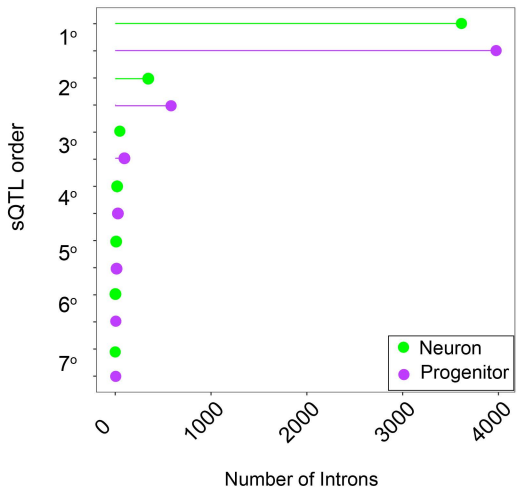
A



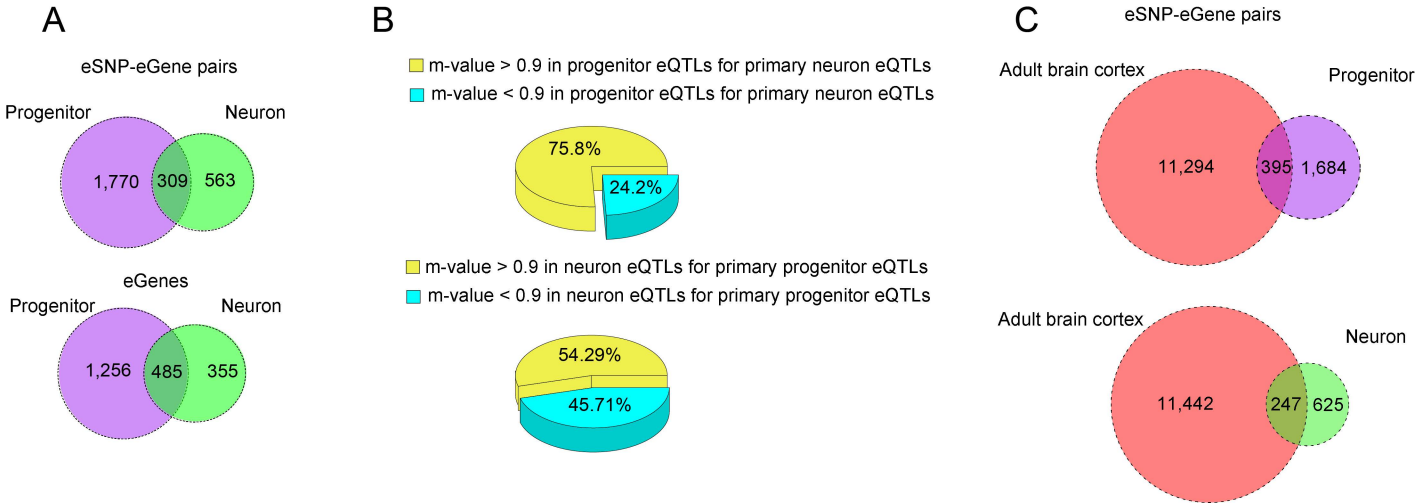
B



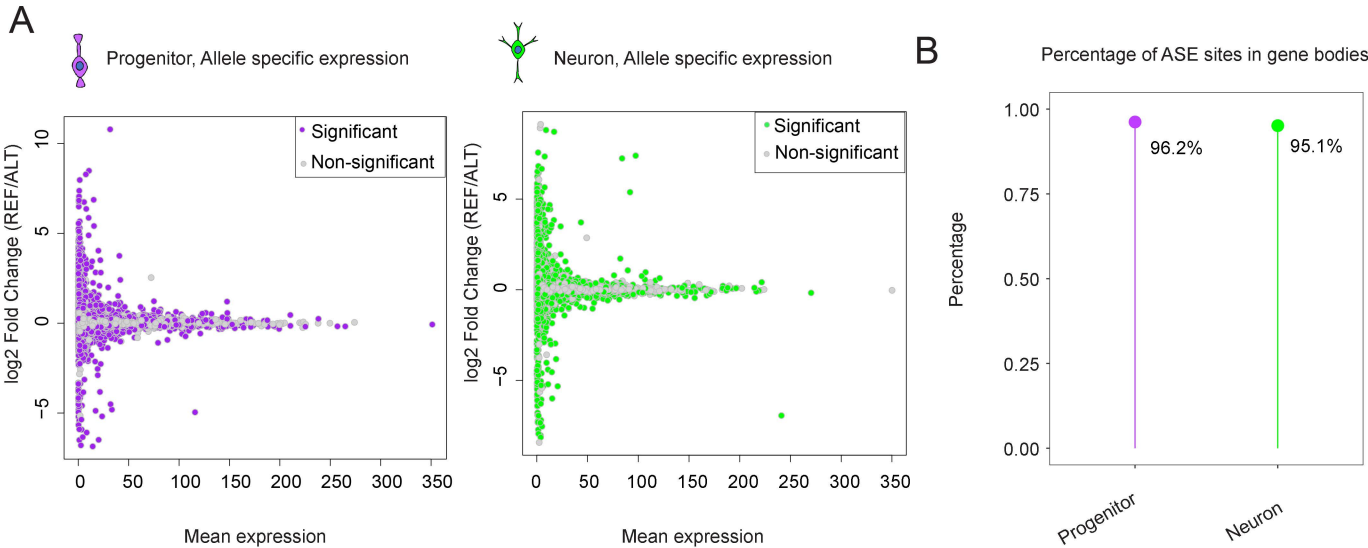
C



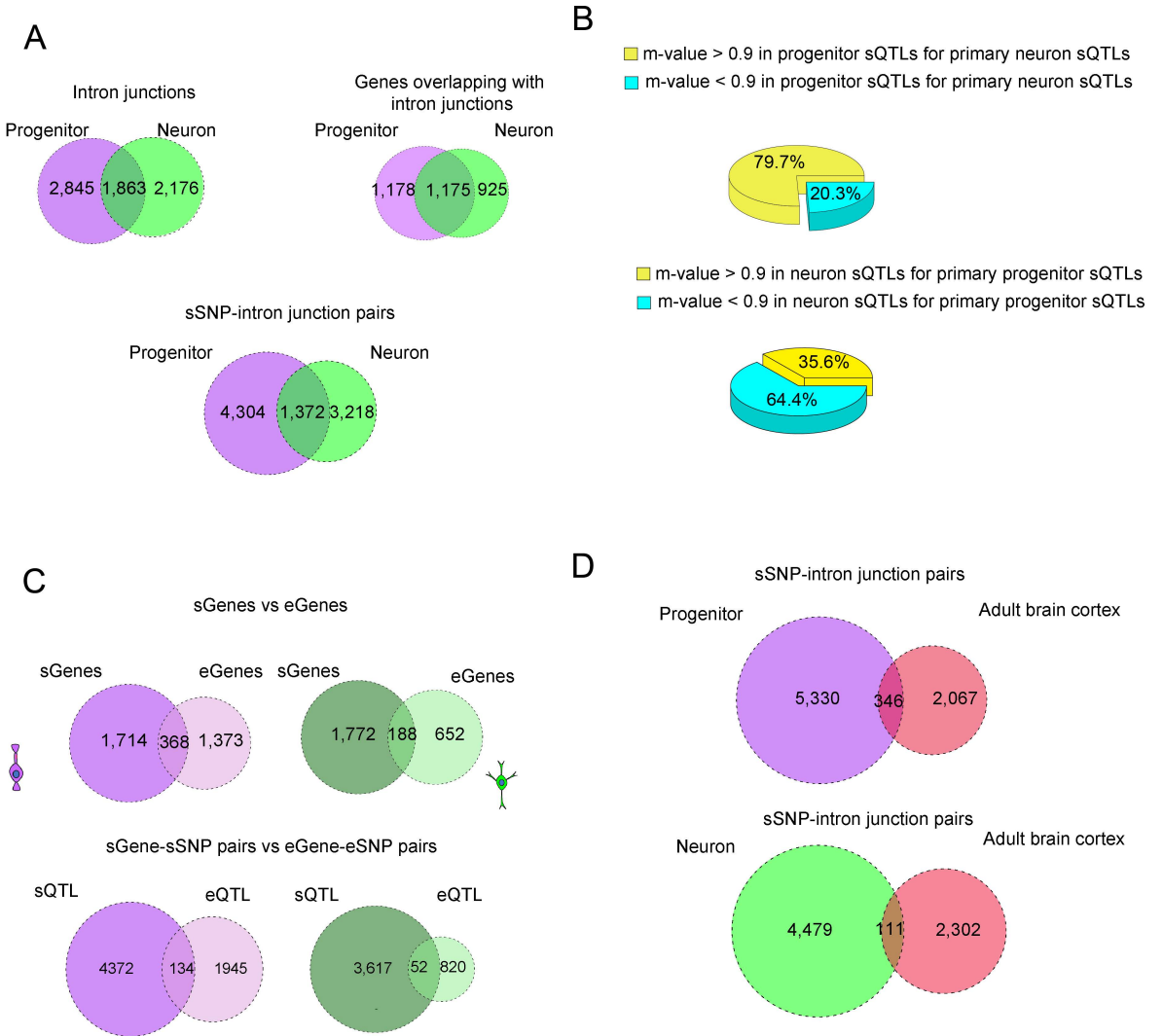
Supplementary Figure 5



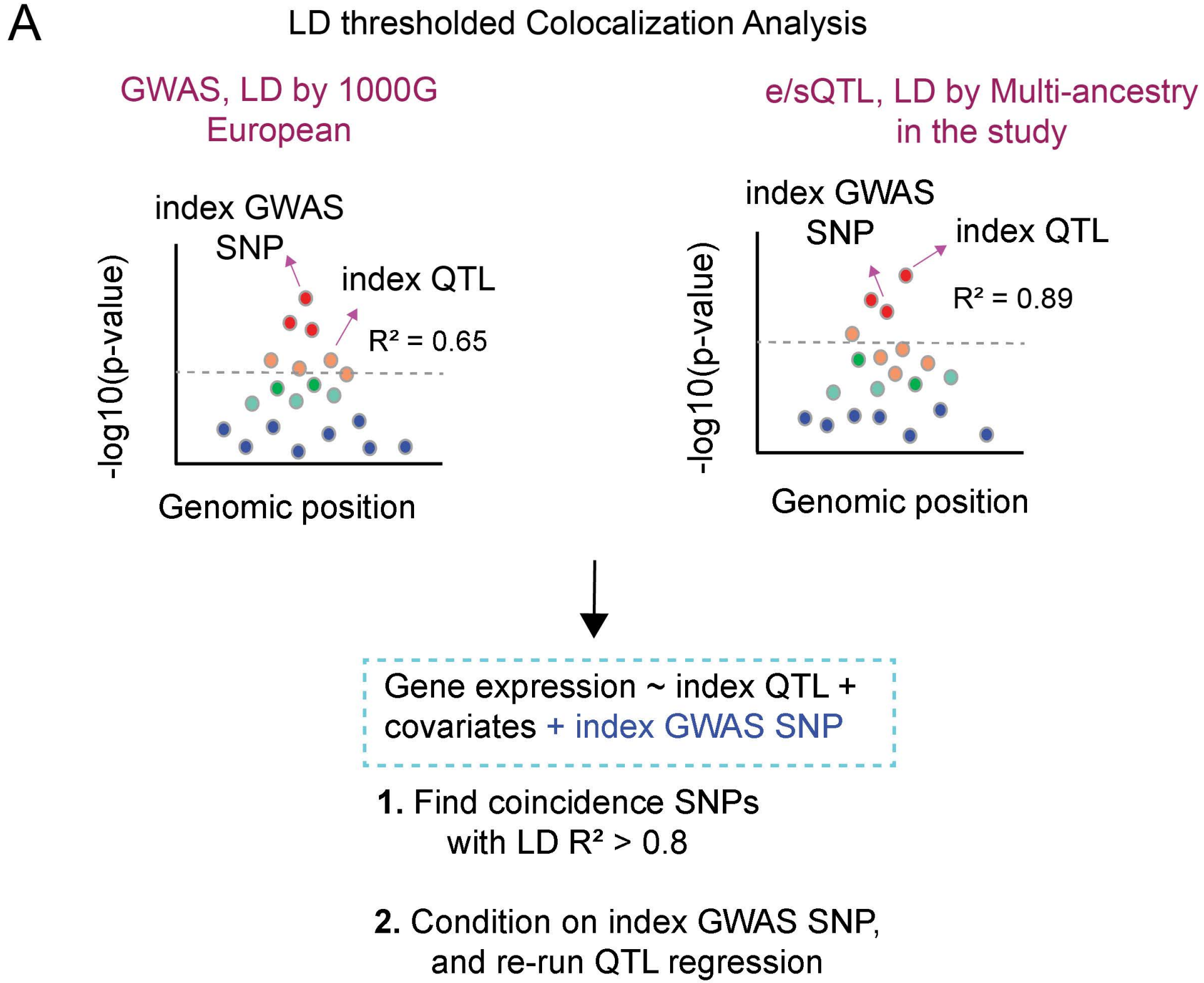
Supplementary Figure 6



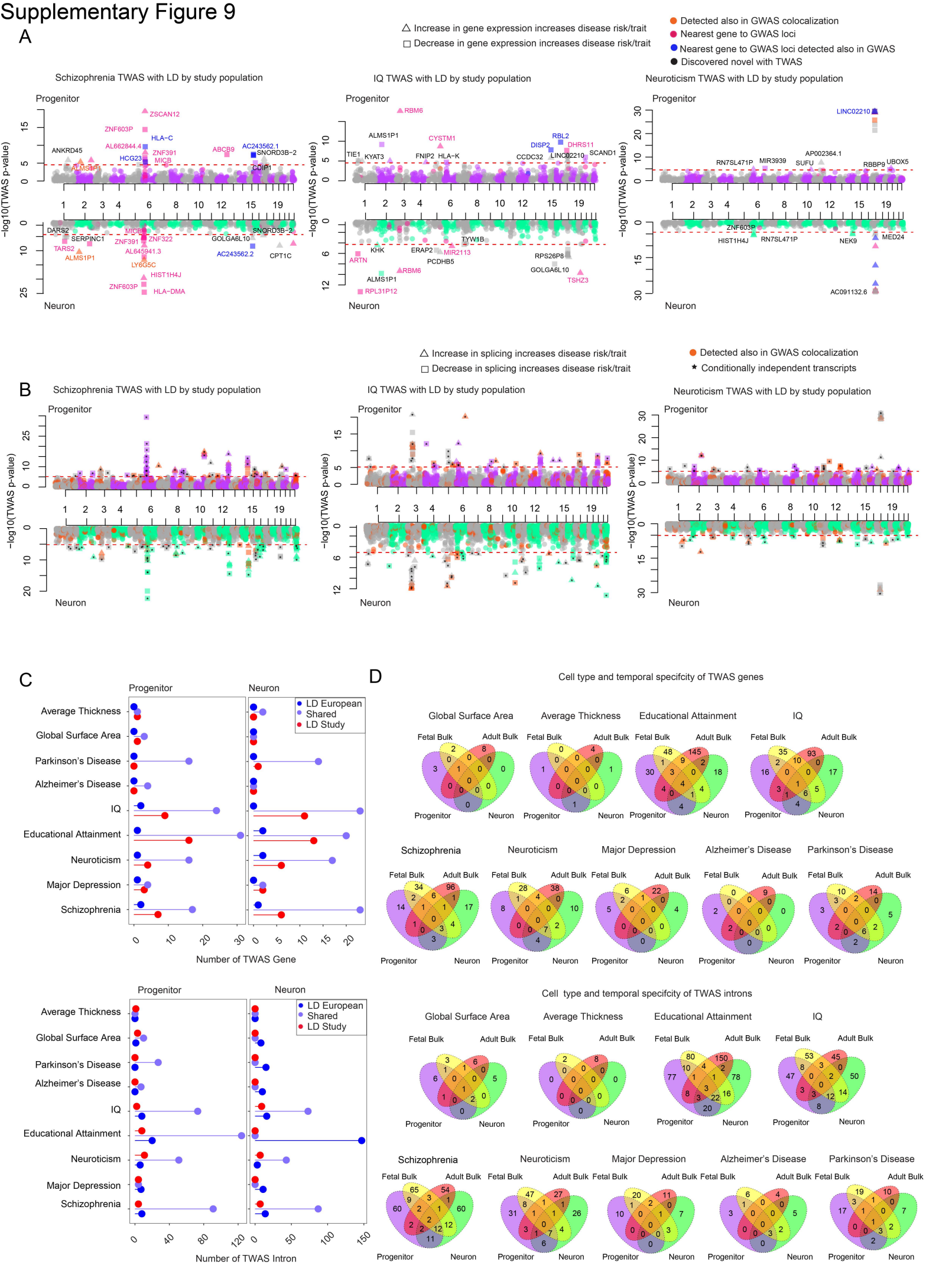
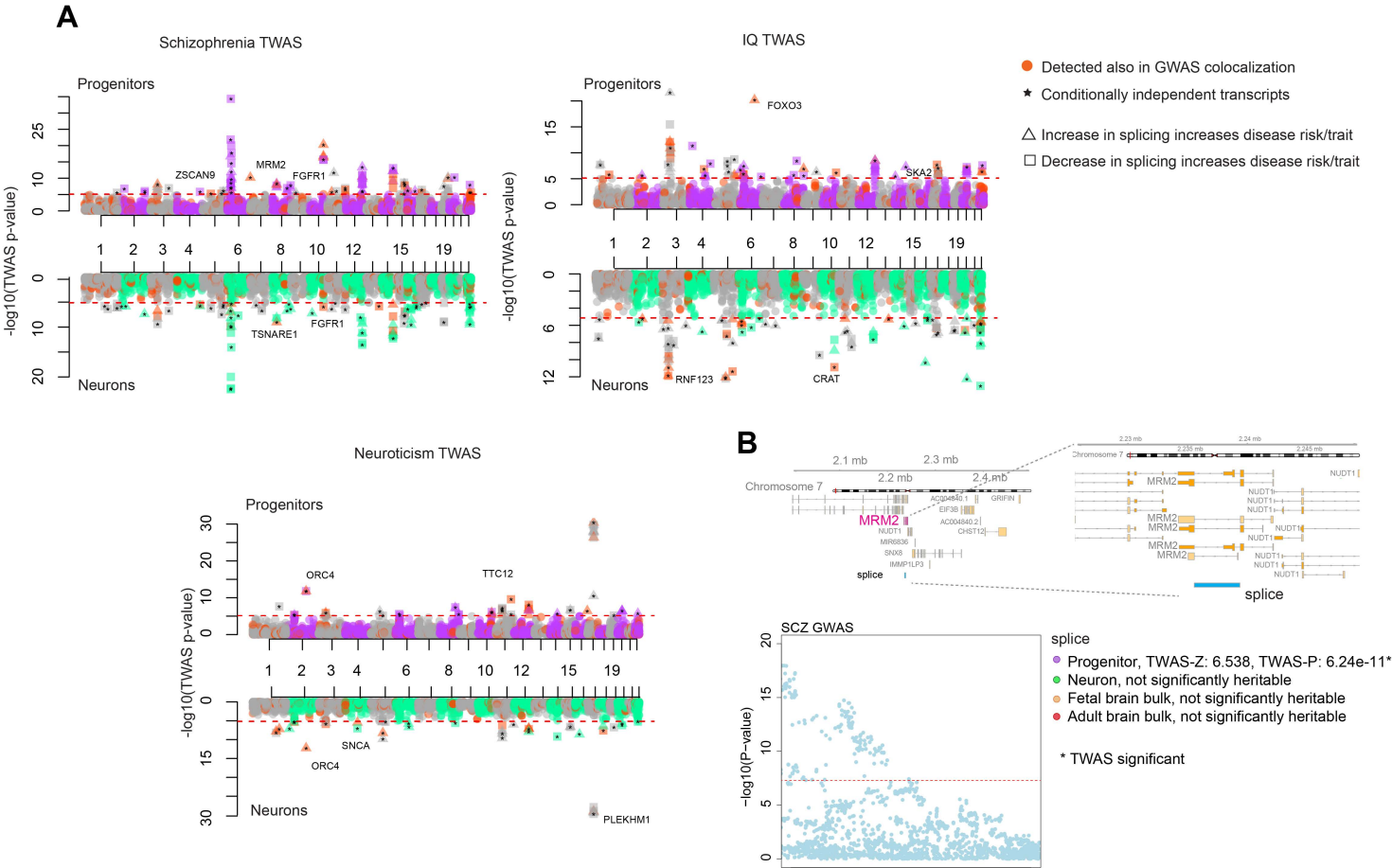
Supplementary Figure 7



Supplementary Figure 8



Supplementary Figure 9



Supplementary Figure 10