The American Journal of Human Genetics, Volume *108*

Supplemental information

Brain-trait-associated variants impact cell-type-specific gene regulation during neurogenesis

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

Figure S1

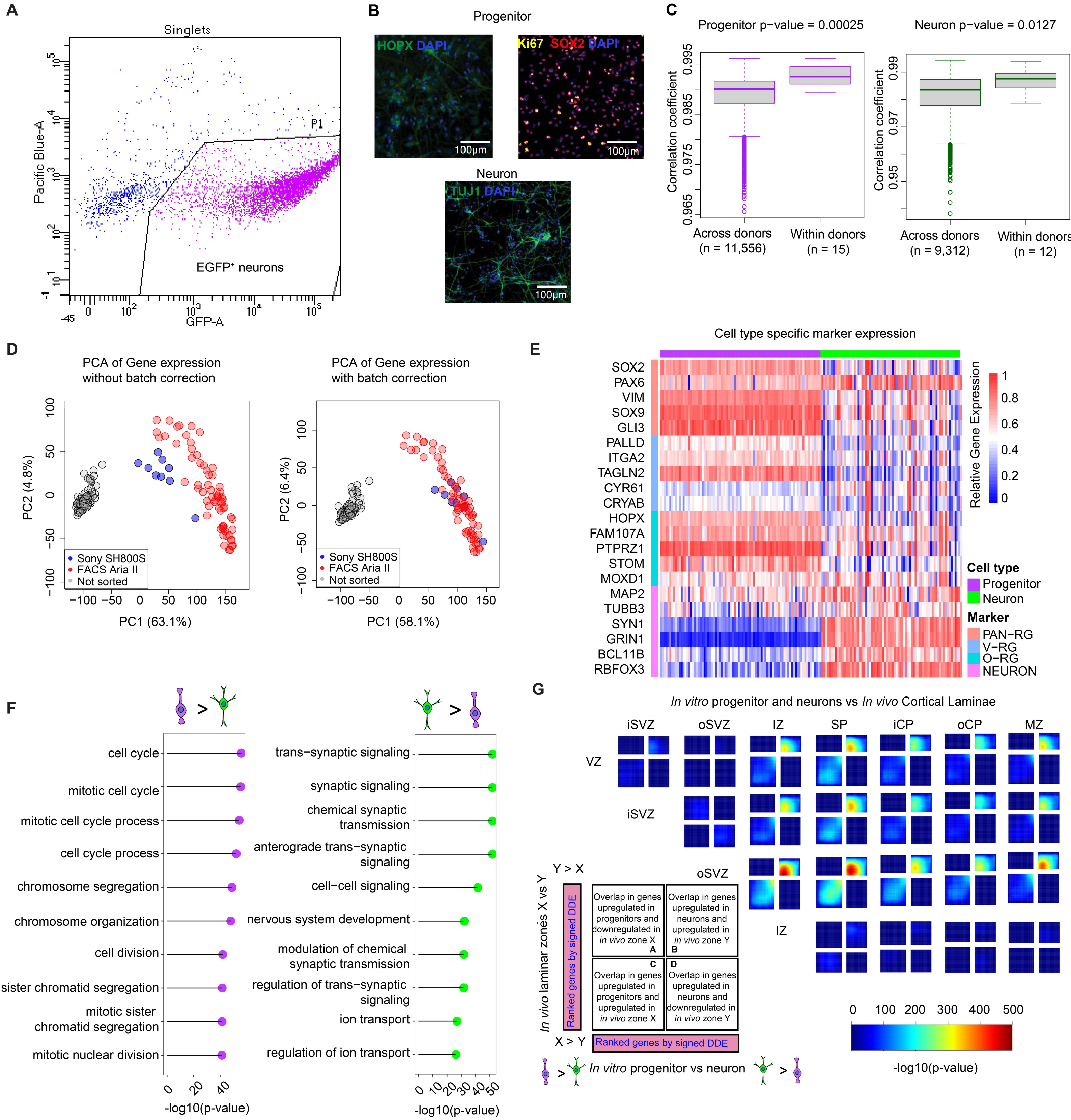

A
Singlets
Pacific Blue-A
GFP-A
P1
EGFP⁺ neurons
B
Progenitor
HOPX DAPI
Ki67 SOX2 DAPI
100µm
Neuron
TUJ1 DAPI
C
Progenitor p-value = 0.00025
Neuron p-value = 0.0127
Correlation coefficient
Across donors (n = 11,556)
Within donors (n = 15)
Across donors (n = 9,312)
Within donors (n = 12)
D
PCA of Gene expression without batch correction
PCA of Gene expression with batch correction
PC2 (4.8%)
PC1 (63.1%)
PC2 (6.4%)
PC1 (58.1%)
Sony SH800S
FACS Aria II
Not sorted
E
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
F
cell cycle
mitotic cell cycle
mitotic cell cycle process
cell cycle process
chromosome segregation
chromosome organization
cell division
sister chromatid segregation
mitotic sister chromatid segregation
mitotic nuclear division
-log10(p-value)
trans−synaptic signaling
synaptic signaling
chemical synaptic transmission
anterograde trans−synaptic signaling
cell−cell signaling
nervous system development
modulation of chemical synaptic transmission
regulation of trans−synaptic signaling
ion transport
regulation of ion transport
-log10(p-value)
G
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 progenitors and downregulated in *in vivo* zone X
A
Overlap in genes upregulated in neurons and upregulated in *in vivo* zone Y
B
C
Overlap in genes upregulated in progenitors and upregulated in *in vivo* zone X
D
Overlap in genes upregulated in neurons and downregulated in *in vivo* zone Y
Ranked genes by signed DDE
In vitro progenitor vs neuron
-log10(p-value)

Figure S1, related to Figure 1: Pre-processing RNA-seq data and evaluation of the fidelity of *in vitro* cell-type specific system.

(A) Flow cytometry results showing sorting of live EGFP positive neurons in pink. The y-axis marks fluorescence from a live/dead stain (annexin V/SYTOX) and the x-axis marks fluorescence from GFP.

(B) Immunolabeling indicates that undifferentiated progenitor cultures were positive for outer radial glia marker HOPX in green, proliferation marker Ki67 in yellow and pan-radial glia marker SOX2 in red, and neurons from 8 week differentiated cultures were positive for the neuronal marker TUJ1 (scale bar is 100 m, DAPI in blue).

(C) 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=0.00025) and neurons (right, p-value=0.0127).

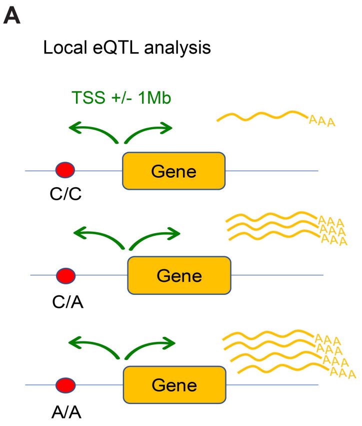
(D) Principal component analysis (PCA) before and after batch correction of neuron for the machine (Sony SH800S in blue, FACS Aria II in red, progenitors not sorted in grey) used for sorting.

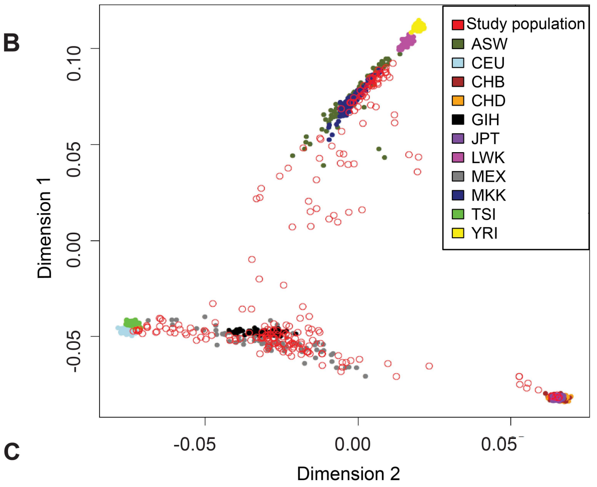
(E) Heatmap showing cell-type specific expression of literature-based progenitor (PAN-RG: Pan-radial glia, V-RG: ventricular radial glia, O-RG: outer radial glia) and neuronal markers listed on the y-axis. The x-axis indicates progenitor (purple) or neuron (green) cells from each donor. The color of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

(F) 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.

(G) 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).

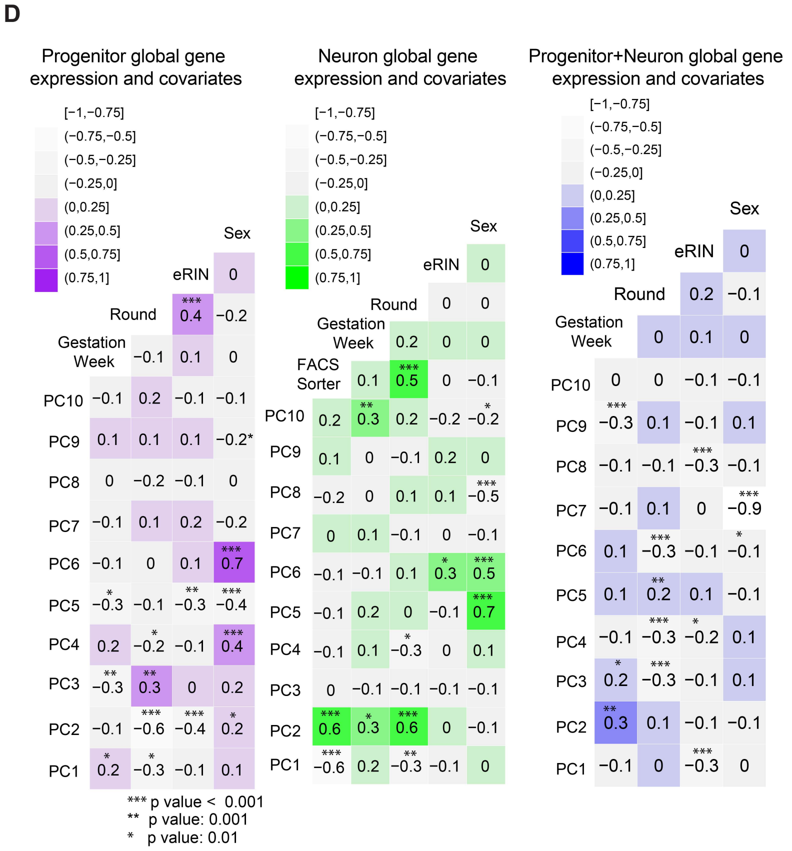
Figure S2





C

	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



Progenitor global gene expression and covariates

	Gestation Week	Round	eRIN	Sex
eRIN				0
Round			*** 0.4	−0.2
Gestation Week		−0.1	0.1	0
PC10	−0.1	0.2	−0.1	−0.1
PC9	0.1	0.1	0.1	−0.2*
PC8	0	−0.2	−0.1	0
PC7	−0.1	0.1	0.2	−0.2
PC6	−0.1	0	0.1	*** 0.7
PC5	* −0.3	−0.1	** −0.3	*** −0.4
PC4	0.2	* −0.2	−0.1	*** 0.4
PC3	** −0.3	** 0.3	0	0.2
PC2	−0.1	*** −0.6	*** −0.4	* 0.2
PC1	* 0.2	* −0.3	−0.1	0.1

*** p value < 0.001
** p value: 0.001
* p value: 0.01

Neuron global gene expression and covariates

	FACS Sorter	Gestation Week	Round	eRIN	Sex
eRIN					0
Round				0	0
Gestation Week			0.2	0	0
FACS Sorter		0.1	*** 0.5	0	−0.1
PC10	0.2	** 0.3	0.2	−0.2	* −0.2
PC9	0.1	0	−0.1	0.2	0
PC8	−0.2	0	0.1	0.1	*** −0.5
PC7	0	0.1	−0.1	0	−0.1
PC6	−0.1	−0.1	0.1	* 0.3	*** 0.5
PC5	−0.1	0.2	0	−0.1	*** 0.7
PC4	−0.1	0.1	* −0.3	0	0.1
PC3	0	−0.1	−0.1	−0.1	−0.1
PC2	*** 0.6	* 0.3	*** 0.6	0	−0.1
PC1	*** −0.6	0.2	** −0.3	−0.1	0

Progenitor+Neuron global gene expression and covariates

	Gestation Week	Round	eRIN	Sex
eRIN				0
Round			0.2	−0.1
Gestation Week		0	0.1	0
PC10	0	0	−0.1	−0.1
PC9	*** −0.3	0.1	−0.1	0.1
PC8	−0.1	−0.1	*** −0.3	−0.1
PC7	−0.1	0.1	0	*** −0.9
PC6	0.1	*** −0.3	−0.1	* −0.1
PC5	0.1	** 0.2	0.1	−0.1
PC4	−0.1	*** −0.3	* −0.2	0.1
PC3	* 0.2	*** −0.3	−0.1	0.1
PC2	** 0.3	0.1	−0.1	−0.1
PC1	−0.1	0	*** −0.3	0

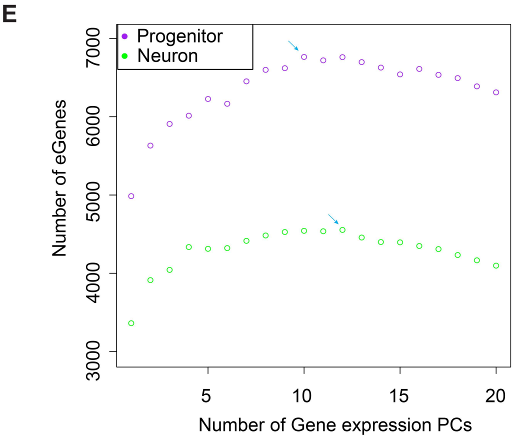


Figure S2, related to Figure 1, Figure 2 and Methods: Local eQTL and the detection of covariates for eQTLs.

(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) 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).

(C) 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.

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

(E) Covariate selection analysis for eQTLs with number of eGene vs. number of global gene expression PCs (progenitors in purple, neurons in green). Blue arrows indicate the number of PCs used in each dataset.

Figure S3

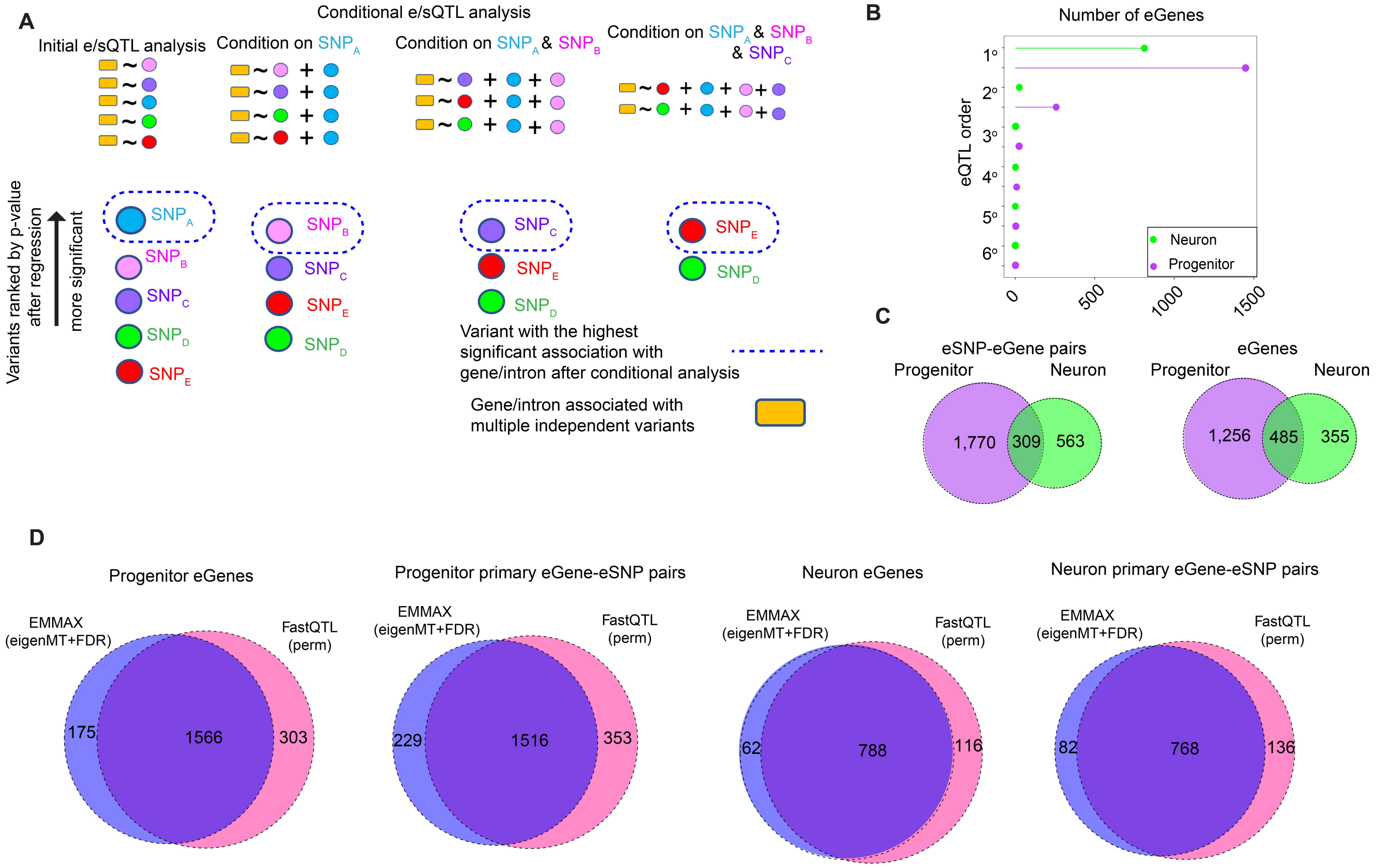

A
Conditional e/sQTL analysis
Initial e/sQTL analysis
Condition on SNP$_A$
Condition on SNP$_A$ & SNP$_B$
Condition on SNP$_A$ & SNP$_B$ & SNP$_C$
Variants ranked by p-value after regression
more significant
SNP$_A$
SNP$_B$
SNP$_C$
SNP$_D$
SNP$_E$
Variant with the highest significant association with gene/intron after conditional analysis
Gene/intron associated with multiple independent variants
B
Number of eGenes
eQTL_order
1°
2°
3°
4°
5°
6°
0
500
1000
1500
Neuron
Progenitor
C
eSNP-eGene pairs
Progenitor
Neuron
1,770
309
563
eGenes
Progenitor
Neuron
1,256
485
355
D
Progenitor eGenes
EMMAX (eigenMT+FDR)
FastQTL (perm)
175
1566
303
Progenitor primary eGene-eSNP pairs
EMMAX (eigenMT+FDR)
FastQTL (perm)
229
1516
353
Neuron eGenes
EMMAX (eigenMT+FDR)
FastQTL (perm)
62
788
116
Neuron primary eGene-eSNP pairs
EMMAX (eigenMT+FDR)
FastQTL (perm)
82
768
136

Figure S3, related to Figure 2 and 3: Conditional QTL analysis and comparison of linear mixed effects models vs standard linear models.

(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 the number of conditionally independent eSNPs on the y-axis indicated by eQTL order (left).

(C) LD-based overlap between progenitor and neuron eQTLs for eSNP-eGene pairs and eGenes.

(D) Comparison of eGenes and primary eGene-eSNP pairs detected by EMMAX followed by eigenMT-FDR and FastQTL followed by adaptive permutation.

Figure S4

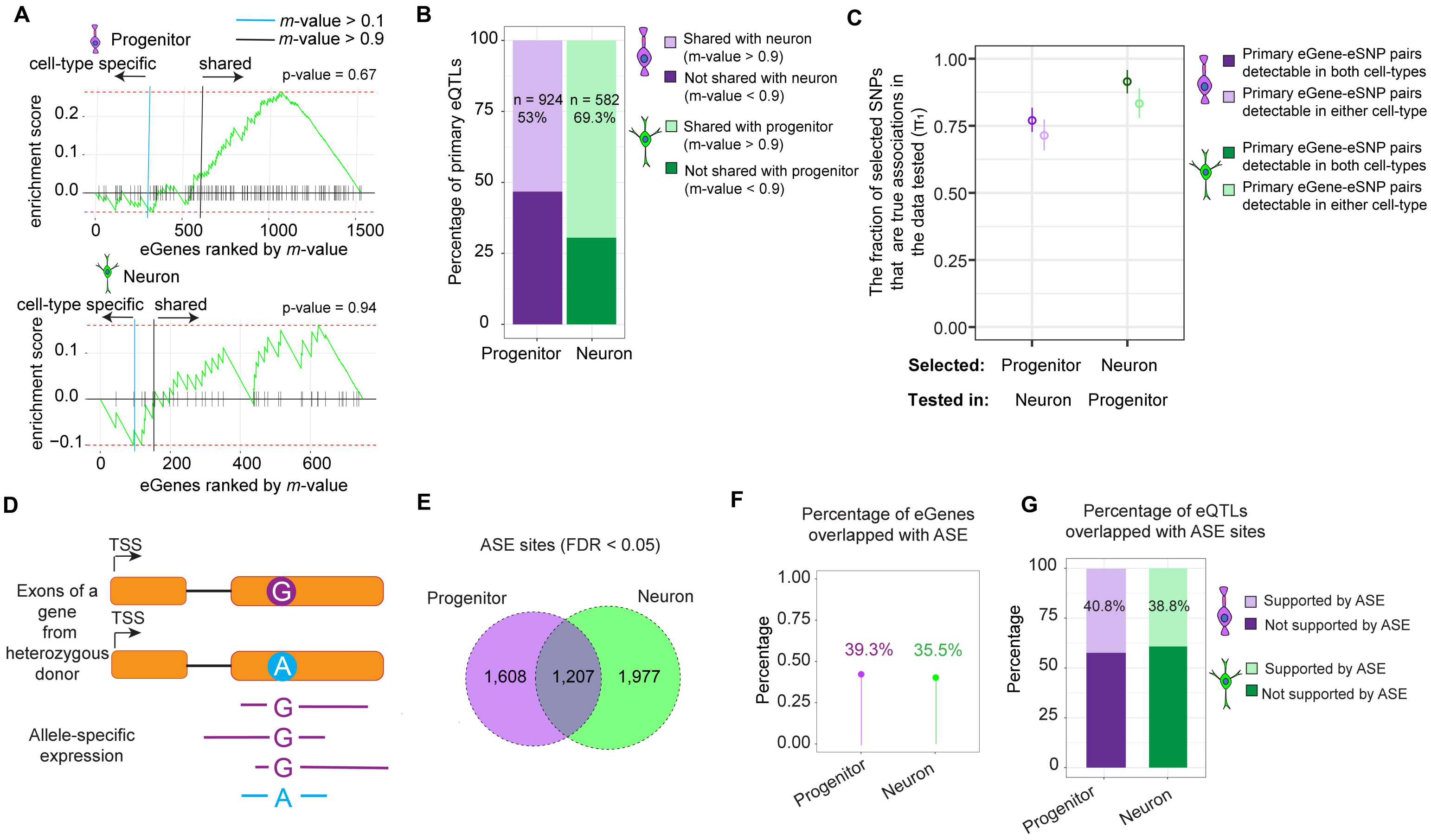

A
Progenitor
m-value > 0.1
m-value > 0.9
cell-type specific
shared
p-value = 0.67
enrichment score
0.2
0.1
0.0
0
500
1000
1500
eGenes ranked by m-value
Neuron
cell-type specific
shared
p-value = 0.94
enrichment score
0.1
0.0
−0.1
0
200
400
600
eGenes ranked by m-value
B
Percentage of primary eQTLs
100
75
50
25
0
n = 924
53%
n = 582
69.3%
Progenitor
Neuron
Shared with neuron (m-value > 0.9)
Not shared with neuron (m-value < 0.9)
Shared with progenitor (m-value > 0.9)
Not shared with progenitor (m-value < 0.9)
C
The fraction of selected SNPs that are true associations in the data tested (π₁)
1.00
0.75
0.50
0.25
0.00
Selected:
Progenitor
Neuron
Tested in:
Neuron
Progenitor
Primary eGene-eSNP pairs detectable in both cell-types
Primary eGene-eSNP pairs detectable in either cell-type
Primary eGene-eSNP pairs detectable in both cell-types
Primary eGene-eSNP pairs detectable in either cell-type
D
TSS
Exons of a gene from heterozygous donor
TSS
G
A
Allele-specific expression
G
G
G
A
E
ASE sites (FDR < 0.05)
Progenitor
Neuron
1,608
1,207
1,977
F
Percentage of eGenes overlapped with ASE
Percentage
1.00
0.75
0.50
0.25
0.00
39.3%
35.5%
Progenitor
Neuron
G
Percentage of eQTLs overlapped with ASE sites
Percentage
100
75
50
25
0
40.8%
38.8%
Progenitor
Neuron
Supported by ASE
Not supported by ASE
Supported by ASE
Not supported by ASE

Figure S4, related to Figure 2: Cross cell-type specific eQTL comparison and ASE analysis.

(A) Gene set enrichment test for cell-type specific eGenes in genes with discordant expression between our *in vitro* culture and the *in vivo* brain. Enrichment p-values are shown. Blue vertical lines indicate genes with *m*-value lower than 0.1 and black vertical lines indicate genes with *m*-value higher than 0.9.

(B) Posterior probability of shared effect size (*m*-value) across cell-types for *m*-value > 0.9.

(C) The fraction of progenitor/neuron primary eGene-eSNP pairs that are true associations (π_1) in neuron/progenitor eQTLs detectable in both cell-types or either cell-types. 95% upper and lower confidence interval are shown.

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

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

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

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

Figure S5

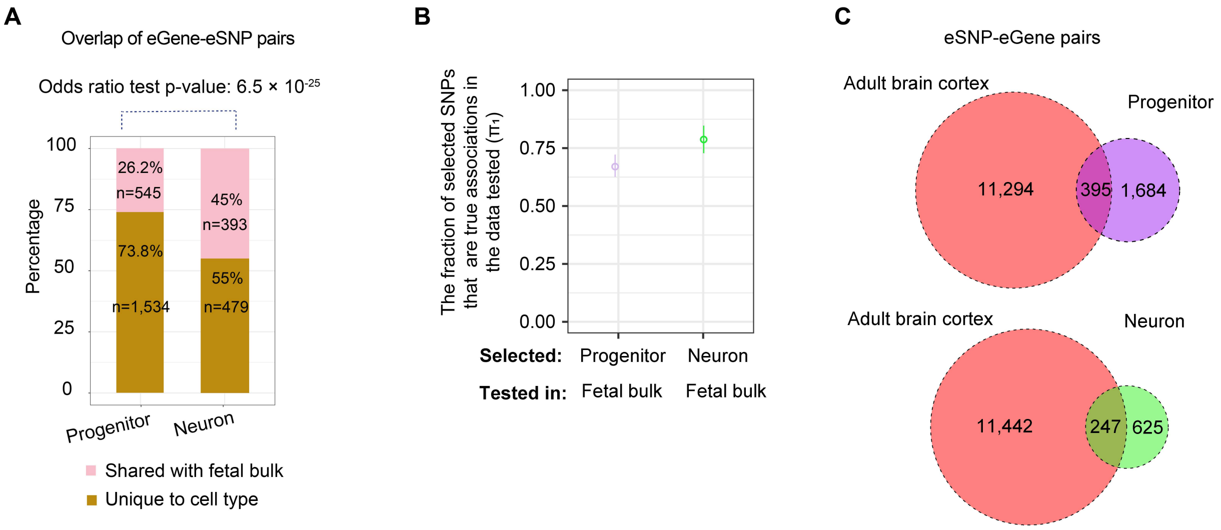

A
Overlap of eGene-eSNP pairs
Odds ratio test p-value: 6.5 × 10-25
100
75
50
25
0
Percentage
26.2%
n=545
73.8%
n=1,534
45%
n=393
55%
n=479
Progenitor
Neuron
Shared with fetal bulk
Unique to cell type
B
The fraction of selected SNPs that are true associations in the data tested (π1)
1.00
0.75
0.50
0.25
0.00
Selected:
Progenitor
Neuron
Tested in:
Fetal bulk
Fetal bulk
C
eSNP-eGene pairs
Adult brain cortex
Progenitor
11,294
395
1,684
Adult brain cortex
Neuron
11,442
247
625

Figure S5, related to Figure 2: Cell-type and temporal specificity of eQTLs.

(**A**) LD-based overlap percentage of cell-type specific eSNP-eGene pairs shared (pink) with fetal bulk eQTLs (variants with LD $r^2 > 0.8$ were considered as the same loci). Odds ratio test p-value is shown.

(B) The fraction of progenitor/neuron primary eGene-eSNP pairs that are true associations (π_1) in fetal bulk eQTLs, subset to genes detectable in either cell-type specific and fetal bulk data. 95% upper and lower confidence interval are shown.

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

Figure S6

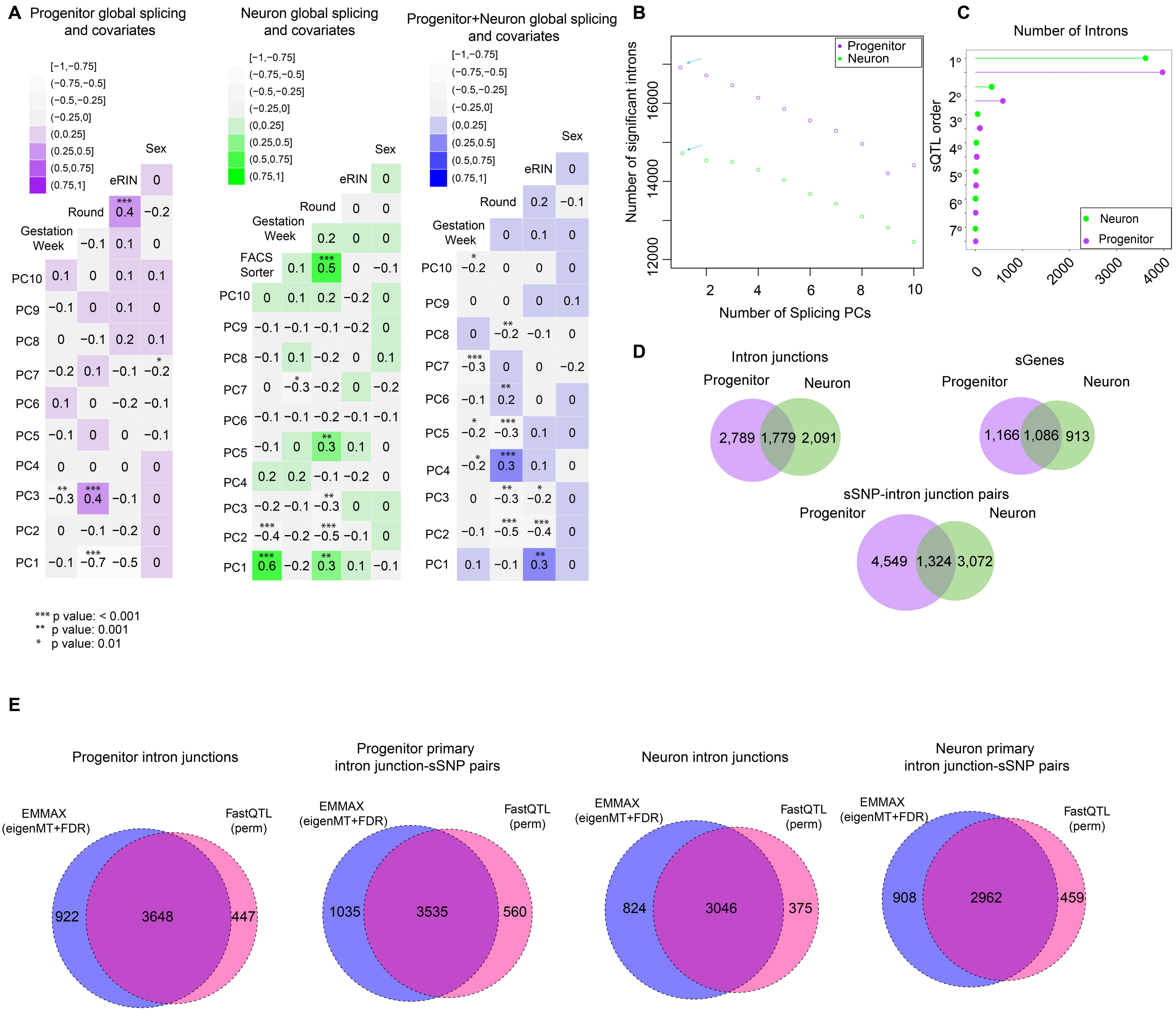

A
Progenitor global splicing and covariates
[−1,−0.75]
(−0.75,−0.5]
(−0.5,−0.25]
(−0.25,0]
(0,0.25]
(0.25,0.5]
(0.5,0.75]
(0.75,1]
Sex
eRIN 0
Round *** 0.4 −0.2
Gestation Week −0.1 0.1 0
PC10 0.1 0 0.1 0.1
PC9 −0.1 0 0.1 0
PC8 0 −0.1 0.2 0.1
PC7 −0.2 0.1 −0.1 * −0.2
PC6 0.1 0 −0.2 −0.1
PC5 −0.1 0 0 −0.1
PC4 0 0 0 0
PC3 ** −0.3 *** 0.4 −0.1 0
PC2 0 −0.1 −0.2 0
PC1 −0.1 *** −0.7 −0.5 0
Neuron global splicing and covariates
[−1,−0.75]
(−0.75,−0.5]
(−0.5,−0.25]
(−0.25,0]
(0,0.25]
(0.25,0.5]
(0.5,0.75]
(0.75,1]
Sex
eRIN 0
Round 0 0
Gestation Week 0.2 0 0
FACS Sorter 0.1 *** 0.5 0 −0.1
PC10 0 0.1 0.2 −0.2 0
PC9 −0.1 −0.1 −0.1 −0.2 0
PC8 −0.1 0.1 −0.2 0 0.1
PC7 0 * −0.3 −0.2 0 −0.2
PC6 −0.1 −0.1 −0.2 −0.1 −0.1
PC5 −0.1 0 ** 0.3 0.1 0
PC4 0.2 0.2 −0.1 −0.2 0
PC3 −0.2 −0.1 ** −0.3 0 0
PC2 *** −0.4 −0.2 *** −0.5 −0.1 0
PC1 *** 0.6 −0.2 ** 0.3 0.1 −0.1
Progenitor+Neuron global splicing and covariates
[−1,−0.75]
(−0.75,−0.5]
(−0.5,−0.25]
(−0.25,0]
(0,0.25]
(0.25,0.5]
(0.5,0.75]
(0.75,1]
Sex
eRIN 0
Round 0.2 −0.1
Gestation Week 0 0.1 0
PC10 * −0.2 0 0 0
PC9 0 0 0 0.1
PC8 0 ** −0.2 −0.1 0
PC7 *** −0.3 0 0 −0.2
PC6 −0.1 ** 0.2 0 0
PC5 * −0.2 *** −0.3 0.1 0
PC4 * −0.2 *** 0.3 0.1 0
PC3 0 ** −0.3 * −0.2 0
PC2 −0.1 *** −0.5 *** −0.4 0
PC1 0.1 −0.1 ** 0.3 0
*** p value: < 0.001
** p value: 0.001
* p value: 0.01
B
Number of significant introns
12000
14000
16000
2
4
6
8
10
Number of Splicing PCs
Progenitor
Neuron
C
Number of Introns
sQTL order
1°
2°
3°
4°
5°
6°
7°
0
1000
2000
3000
4000
Neuron
Progenitor
D
Intron junctions
Progenitor
Neuron
2,789
1,779
2,091
sGenes
Progenitor
Neuron
1,166
1,086
913
sSNP-intron junction pairs
Progenitor
Neuron
4,549
1,324
3,072
E
Progenitor intron junctions
EMMAX (eigenMT+FDR)
FastQTL (perm)
922
3648
447
Progenitor primary intron junction-sSNP pairs
EMMAX (eigenMT+FDR)
FastQTL (perm)
1035
3535
560
Neuron intron junctions
EMMAX (eigenMT+FDR)
FastQTL (perm)
824
3046
375
Neuron primary intron junction-sSNP pairs
EMMAX (eigenMT+FDR)
FastQTL (perm)
908
2962
459

Figure S6, related to Figure 3: Cell-type specific sQTL and comparison of linear mixed effects models vs standard linear models

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

(B) Covariate selection analysis 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.

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

(D) LD-based overlap of intron junctions, sGenes harboring intron junctions, and sSNP-intron junction pairs for progenitor vs neuron sQTLs.

(E) Comparison of intron junctions and primary intron junction-sSNP pairs detected by EMMAX followed by eigenMT-FDR and FastQTL followed by adaptive permutation.

Figure S7

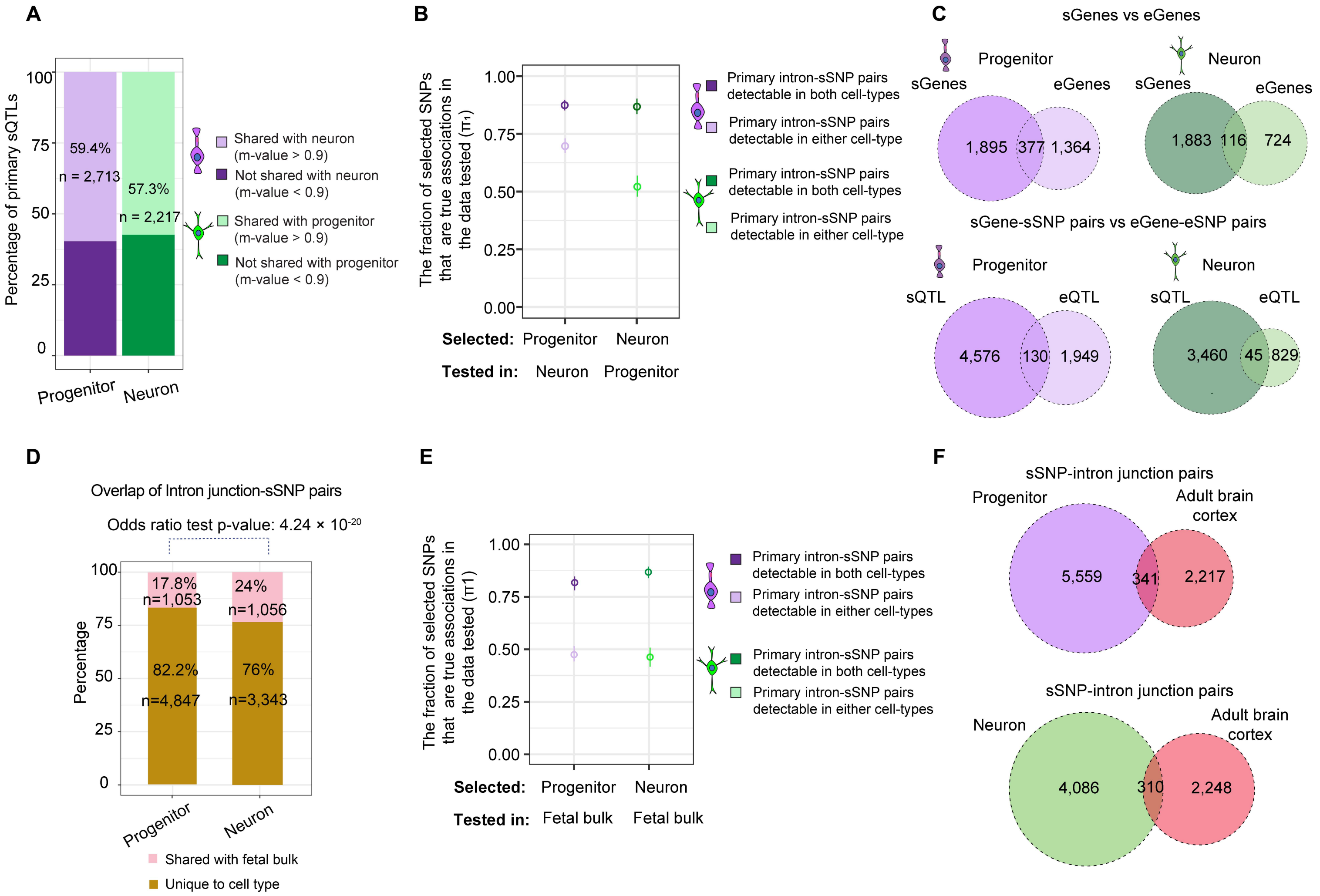

A
Percentage of primary sQTLs
100
75
50
25
0
59.4%
n = 2,713
57.3%
n = 2,217
Progenitor
Neuron
Shared with neuron (m-value > 0.9)
Not shared with neuron (m-value < 0.9)
Shared with progenitor (m-value > 0.9)
Not shared with progenitor (m-value < 0.9)
B
The fraction of selected SNPs that are true associations in the data tested (π₁)
1.00
0.75
0.50
0.25
0.00
Selected: Progenitor Neuron
Tested in: Neuron Progenitor
Primary intron-sSNP pairs detectable in both cell-types
Primary intron-sSNP pairs detectable in either cell-type
Primary intron-sSNP pairs detectable in both cell-types
Primary intron-sSNP pairs detectable in either cell-type
C
sGenes vs eGenes
Progenitor
sGenes
eGenes
1,895
377
1,364
Neuron
sGenes
eGenes
1,883
116
724
sGene-sSNP pairs vs eGene-eSNP pairs
Progenitor
sQTL
eQTL
4,576
130
1,949
Neuron
sQTL
eQTL
3,460
45
829
D
Overlap of Intron junction-sSNP pairs
Odds ratio test p-value: 4.24 × 10⁻²⁰
Percentage
100
75
50
25
0
17.8%
n=1,053
24%
n=1,056
82.2%
n=4,847
76%
n=3,343
Progenitor
Neuron
Shared with fetal bulk
Unique to cell type
E
The fraction of selected SNPs that are true associations in the data tested (π1)
1.00
0.75
0.50
0.25
0.00
Selected: Progenitor Neuron
Tested in: Fetal bulk Fetal bulk
Primary intron-sSNP pairs detectable in both cell-types
Primary intron-sSNP pairs detectable in either cell-types
Primary intron-sSNP pairs detectable in both cell-types
Primary intron-sSNP pairs detectable in either cell-types
F
sSNP-intron junction pairs
Progenitor
Adult brain cortex
5,559
341
2,217
sSNP-intron junction pairs
Neuron
Adult brain cortex
4,086
310
2,248

Figure S7, related to Figure 3: Cell-type and temporal specificity of sQTLs.

(A) Posterior probability of shared effects across cell-type specific sQTLs with *m*-value > 0.9.

(B) The fraction of progenitor/neuron primary intron-sSNP pairs that are true associations (π_1) in neuron/progenitor sQTLs detectable in both or either cell-types. 95% upper and lower confidence interval are shown.

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

(D) LD-based overlap percentage of cell-type specific sSNP-intron junction pairs shared (pink) with fetal bulk sQTLs (variants with LD r^2 > 0.8 were considered as the same loci). Odds ratio test p-value is shown.

(E) The fraction of progenitor/neuron primary intron-sSNP pairs that are true associations (π_1) in fetal bulk sQTLs detectable in both or either cell-type specific and fetal bulk data. 95% upper and lower confidence interval are shown.

(F) LD-based overlap between progenitor (in purple)/neuron (in green) sQTLs and adult brain cortex sQTLs (in red) for intron junction-sSNP pairs.

Figure S8

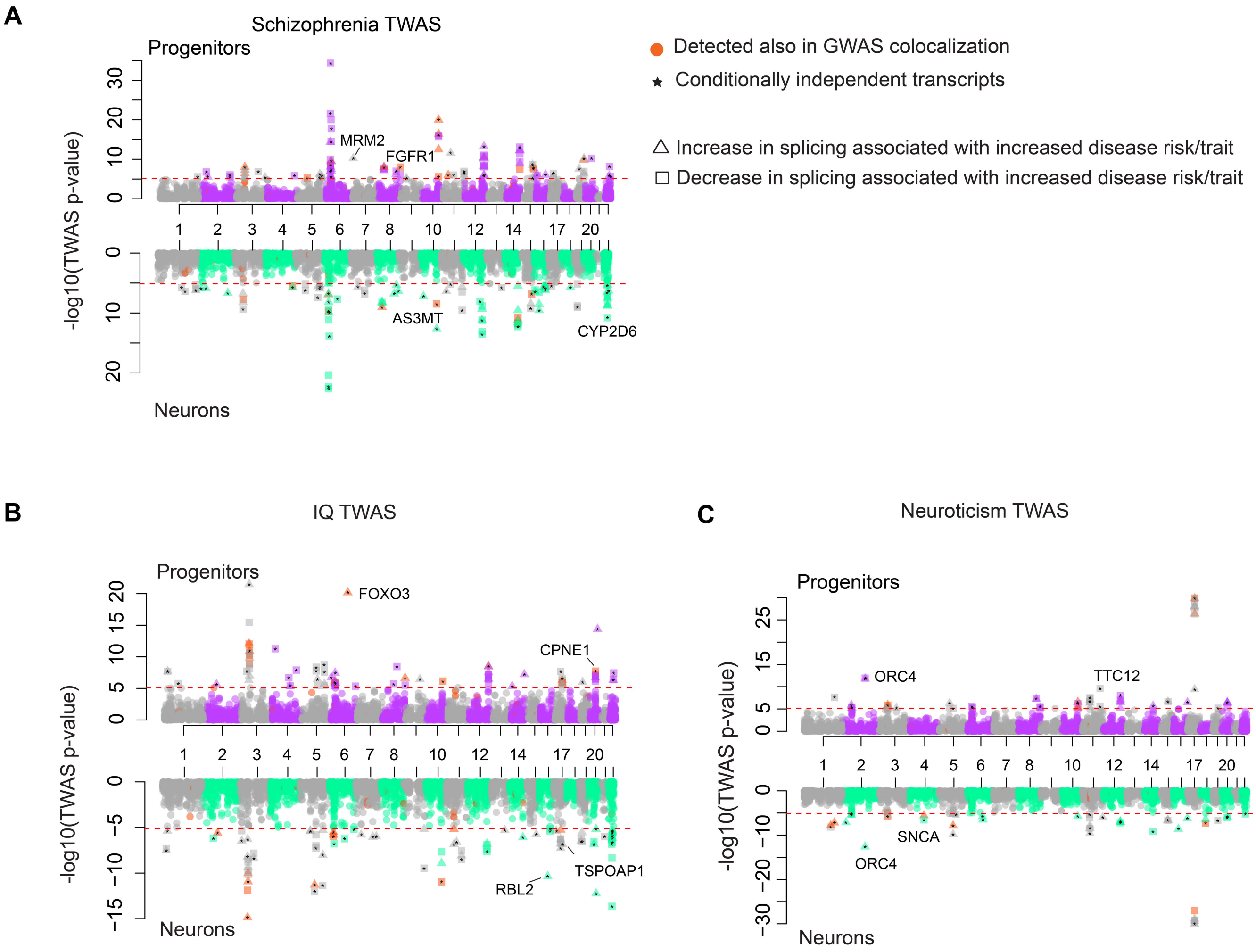

A
Schizophrenia TWAS
Progenitors
Neurons
-log10(TWAS p-value)
MRM2
FGFR1
AS3MT
CYP2D6
● Detected also in GWAS colocalization
★ Conditionally independent transcripts
△ Increase in splicing associated with increased disease risk/trait
□ Decrease in splicing associated with increased disease risk/trait
B
IQ TWAS
Progenitors
Neurons
-log10(TWAS p-value)
FOXO3
CPNE1
TSPOAP1
RBL2
C
Neuroticism TWAS
Progenitors
Neurons
-log10(TWAS p-value)
ORC4
TTC12
SNCA
ORC4
1
2
3
4
5
6
7
8
10
12
14
17
20

Figure S8, related to Figure 6: Prediction of differential alternative splicing events during human brain development via TWAS.

(A) Manhattan plots for schizophrenia TWAS for progenitors (purple-grey, top) and neurons (green-grey, bottom) where LD matrix was 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 (asterisk), where positively and negatively correlated splicing represented by triangle and square, respectively.

(**B**) Manhattan plots for IQ TWAS with graphic design described in A.

(**C**) Manhattan plots for Neuroticism TWAS with graphic design described in A.

Figure S9

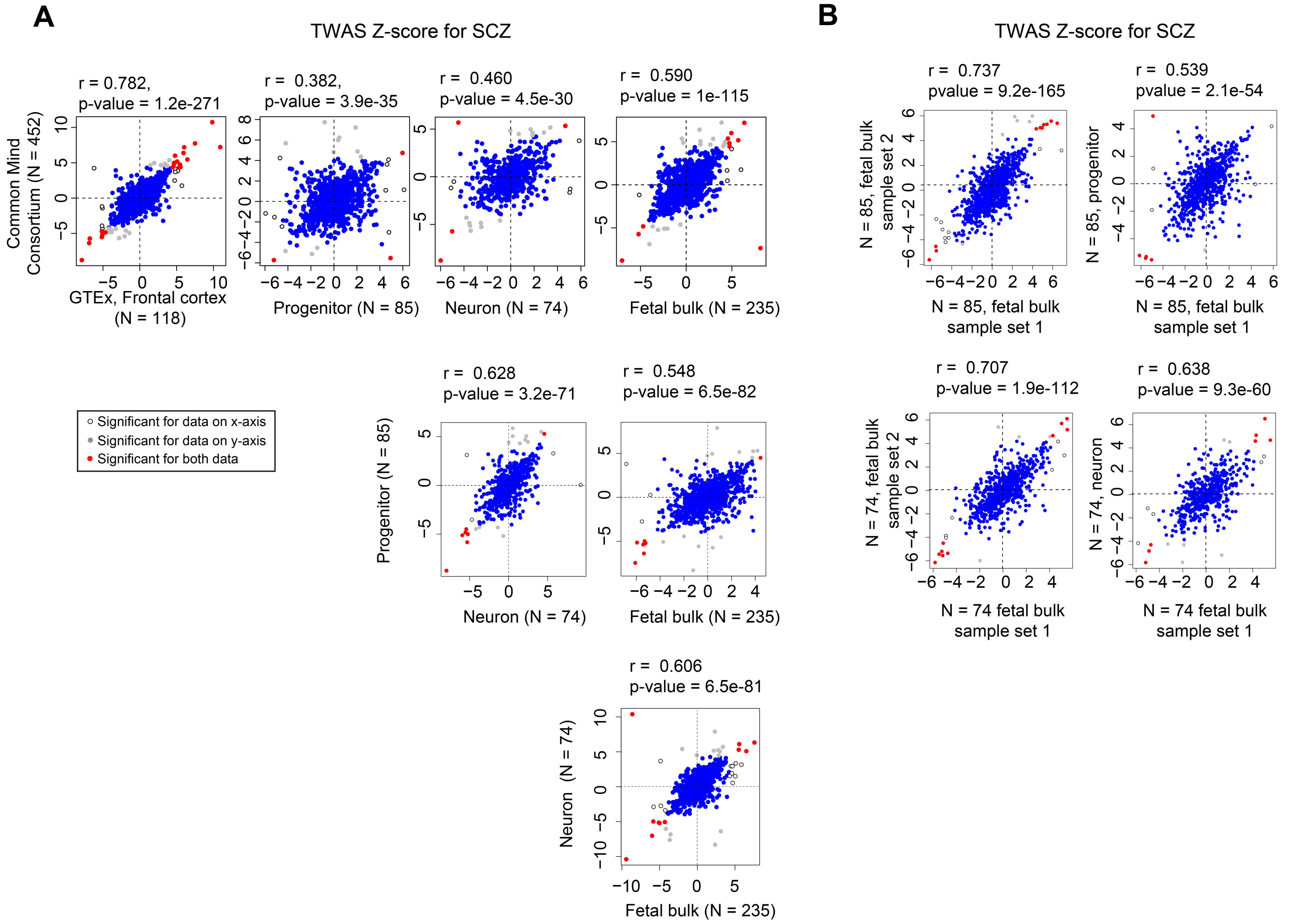

A
TWAS Z-score for SCZ
r = 0.782,
p-value = 1.2e-271
Common Mind Consortium (N = 452)
GTEx, Frontal cortex (N = 118)
r = 0.382,
p-value = 3.9e-35
Progenitor (N = 85)
r = 0.460
p-value = 4.5e-30
Neuron (N = 74)
r = 0.590
p-value = 1e-115
Fetal bulk (N = 235)
Significant for data on x-axis
Significant for data on y-axis
Significant for both data
r = 0.628
p-value = 3.2e-71
Progenitor (N = 85)
Neuron (N = 74)
r = 0.548
p-value = 6.5e-82
Fetal bulk (N = 235)
r = 0.606
p-value = 6.5e-81
Neuron (N = 74)
Fetal bulk (N = 235)
B
TWAS Z-score for SCZ
r = 0.737
pvalue = 9.2e-165
N = 85, fetal bulk sample set 2
N = 85, fetal bulk sample set 1
r = 0.539
pvalue = 2.1e-54
N = 85, progenitor
N = 85, fetal bulk sample set 1
r = 0.707
p-value = 1.9e-112
N = 74, fetal bulk sample set 2
N = 74 fetal bulk sample set 1
r = 0.638
p-value = 9.3e-60
N = 74, neuron
N = 74 fetal bulk sample set 1

Figure S9, related to Figure 6: Evaluation of the impact of sample size on TWAS results.

(A) Comparison of TWAS Z-score for SCZ performed either with CMC adult brain eQTL data (N = 452) and with GTEx adult brain eQTL data (N = 118), or cell-type specific and fetal bulk eQTL data. The genes that were significant for using both datasets are colored in red, for the data on the x-axis shown in white, and for the data on the y-axis in grey.

(B) Comparison of TWAS Z-score for SCZ performed either with fetal bulk eQTL data downsampled for progenitor eQTL sample size (N = 85), or for neuron eQTL sample size (N = 74).

Figure S10

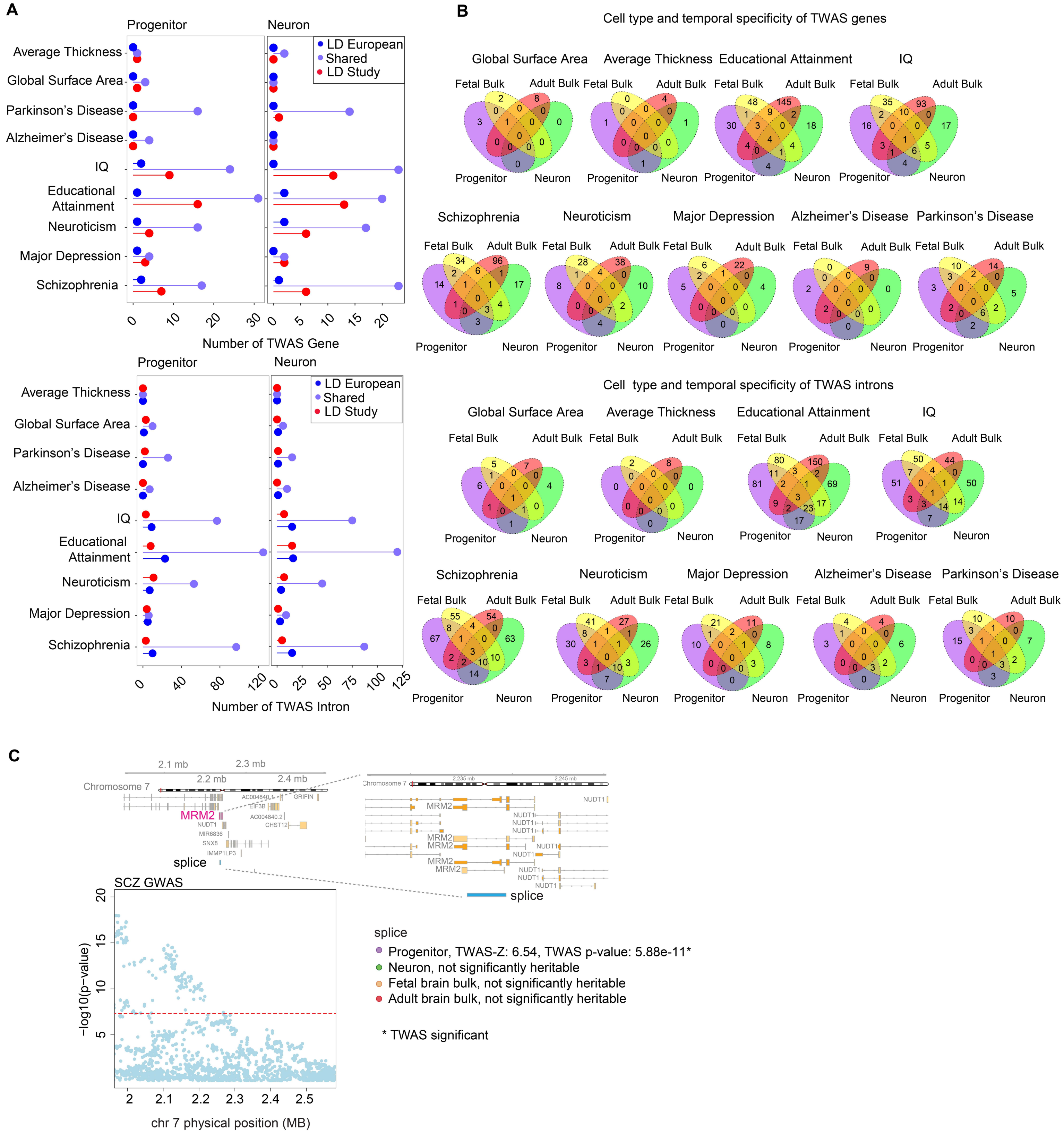

A
Progenitor
Neuron
LD European
Shared
LD Study
Average Thickness
Global Surface Area
Parkinson's Disease
Alzheimer's Disease
IQ
Educational Attainment
Neuroticism
Major Depression
Schizophrenia
Number of TWAS Gene
Number of TWAS Intron
B
Cell type and temporal specificity of TWAS genes
Cell type and temporal specificity of TWAS introns
Fetal Bulk
Adult Bulk
Progenitor
Neuron
C
Chromosome 7
2.1 mb
2.2 mb
2.3 mb
2.4 mb
MRM2
NUDT1
splice
SCZ GWAS
−log10(p−value)
chr 7 physical position (MB)
splice
Progenitor, TWAS-Z: 6.54, TWAS p-value: 5.88e-11*
Neuron, not significantly heritable
Fetal brain bulk, not significantly heritable
Adult brain bulk, not significantly heritable
* TWAS significant

Figure S10, related to Figure 6: Cell-type/temporal specificity of TWAS genes and introns.

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

(**B**) Overlap of cell-type specific TWAS genes (from the analysis where LD was estimated from European population) with fetal brain bulk and adult brain bulk TWAS genes (upper plot). Overlap of cell-type specific TWAS introns (from the analysis where LD was estimated from European population) with fetal brain bulk and adult brain bulk TWAS introns (lower plot). (**C**) 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 x 10^{-8}).

Supplementary Table legends

Table S1, related to Figure 1-2 and S1:

Sheet 1: 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 [37]. 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, related to Figure 2 and S2-4:

Sheet 1-3: 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; pvalue 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 (NA if rsid is not available for the genomic position of the variant in 1000 Genome data; * if multiple variants exist for the same genomic position).

Sheet 4-5: 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[36]; 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 S3, related to Figure 3 and S6-7:

Sheet 1: 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[37]. 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; gene is the gene symbol of the gene that introns junctions overlap with; ensemblID is the ensemblID of that gene.

Sheet 2-4: 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.

Sheet 5: Enrichment of RNA binding protein (RBP) 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[50]; 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; Celltype is the cell type used for enrichment test.

Table S4, related to Figure 4 and 5: 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; 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; trait is the GWAS study; A1 is the effect allele for e/sQTL index SNP; GWASsnp is the variant e/sSNP colocalized with; rsid is the rs id of the allele matching in 1000 Genome Phase 3.

Table S5, related to Figure 6, S8-10:

Sheet 1-8: List of cell-type specific/fetal bulk/adult bulk TWAS gene and introns for neuropsychiatric disease and other brain related traits. Output from FUSION[79]: 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).

Sheet 9-10: Summary of heritability ($p\text{-value} < 0.01$) 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 e/sQTL Z-scores computation with p-value as top1.pval. 95 % confidence intervals per parameter are shown their below.

Sheet 11-12: SCZ TWAS for GTEx Brain frontal cortex and downsampled fetal bulk data. Output from FUSION[78]: PANEL: Data type; 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.