Supplementary document

Brain transcriptomic profiling reveals common patterns across neurodegenerative and psychiatric disorders

Iman Sadeghi, Juan D. Gispert, Emilio Palumbo, Manuel Muñoz-Aguirre, Valentin Wucher, Valeria D'Argenio, Gabriel Santpere, Arcadi Navarro, Roderic Guigo*, Natàlia Vilor-Tejedor*

Contents:

Table S1. RNA-Seq expression datasets used in this study.

Diagnosis	Study/Project	Accession ID	# Samples		Brain region	Ref
			Controls	Cases		
AD	Wang et al	syn3159438	280	742	Anterior prefrontal cortex, perirhinal cortex, superior temporal gyrus, pars opercularis	[1]
	Allen et al	syn5550404	155	164	Temporal cortex, cerebellum	[2]
PD	Dumitriu et al	PRJNA283498	44	29	BA9	[3]
PA	Allen et al	syn5550404	Mentioned (155)	58	Temporal cortex, cerebellum	[2]
PSP	Allen et al	syn5550404	Mentioned (155)	168	Temporal cortex, cerebellum	[2]
Scz	Jaffe et al	syn12299750	320	175	Dorsolateral prefrontal cortex (DLPFC)	[4]
	Xiao et al	PRJNA235930	12	10	BA9, BA24	[5]
	Chang et al	PRJNA379666	24	22	Amygdala	[6]
	Corley et al	PRJNA343829	19	19	DLPFC	[7]
	Wu et al	PRJEB2939	9	9	Superior temporal gyrus	[8]
	CommonMind consortium	syn18097439	294	47	DLPFC	[9]
	CommonMind consortium-HBCC	syn18097439	167	87	DLPFC	[9]
	BrainGVEX	syn4590909	257	95	Frontal cortex	[10]
	Ramaker et al	PRJNA319583	70	71	anterior cingulate cortex, DLPFC,	[11]

					nucleus accumbens	
ASD	Wright et al	PRJNA398545	39	13	DLPFC	[12]
	Li J et al	PRJNA263196	6	6	Corpus callosum	[13]
	Liu et al	PRJNA254971	38	34	Superior frontal gyrus	[14]
	Yale-ASD	syn4566141	30	15	DLPFC	[10]
	UCLA-ASD	syn4587609	126	119	BA4/6, BA38, BA7, BA17	[10]
MDD	Ramaker et al	PRJNA319583	Mentioned (70)	69	anterior cingulate cortex, DLPFC, nucleus accumbens	[11]
	Pantazatos et al	PRJNA394722	29	30	DLPFC (BA9)	[15]
	Labonte et al	PRJNA398031	122	141	Orbitofrontal cortex (BA11), DLPFC (BA8/9), cingulate gyrus 25 (BA25; vmPFC), anterior insula, nucleus accumbens, ventral subiculum	[16]
BP	Xiao et al	PRJNA235930	Mentioned (12)	14	BA9, BA24	[5]
	CommonMind consortium	syn18097439	Mentioned (294)	257	DLPFC	[9]
	CommonMind consortium-HBCC	syn18097439	Mentioned (167)	58	DLPFC	[9]
	BrainGVEX	syn4590909	Mentioned (257)	73	Frontal cortex	[10]
	Ramaker et al	PRJNA319583	Mentioned (70)	71	anterior cingulate cortex, DLPFC, nucleus accumbens	[11]
	Pacifico et al	PRJNA318642	18	18	Dorsal striatum	[17]

	McMullen et al	PRJNA321439	8	8	Putamen, Caudate nucleus	[18]
	Akula et al	PRJNA231202	11	11	DLPFC	[19]

Table S2. The number of studies per region/condition

	cerebellum	Temporal	Frontal	Limbic	Occipital	Insular	Basal ganglia
AD	1	2	1				
PD			1				
PA	1	1					
PSP	1	1					
Scz		2	8	2			1
ASD	2	1	4	1	1		
MDD			3	2		1	2
BP			5	2			3

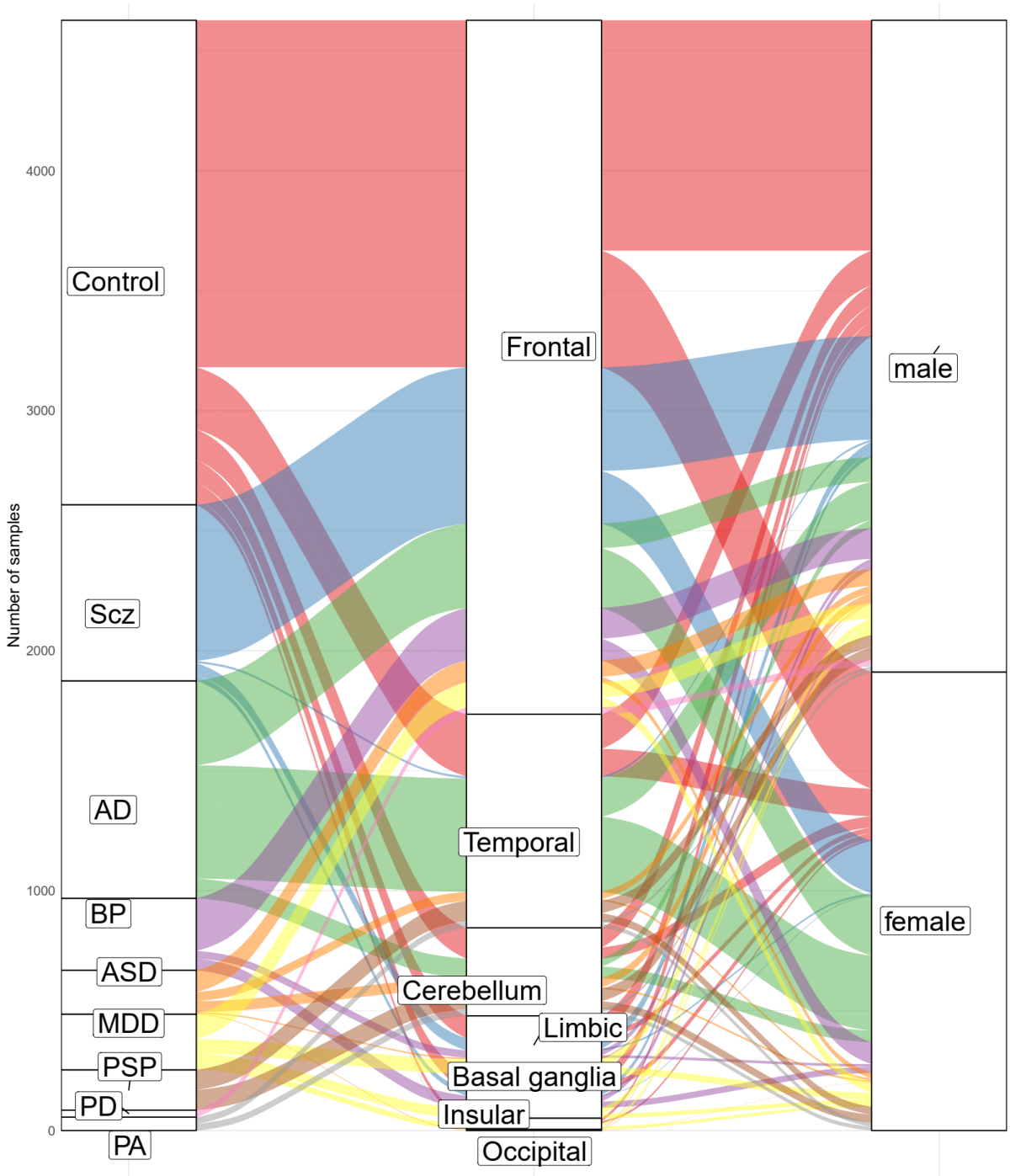

Figure S1. The overview of samples used in this study. A flowchart of the samples. The panels represent diagnosis, brain lobes, and sex from left to right, with colors showing the diagnosis.

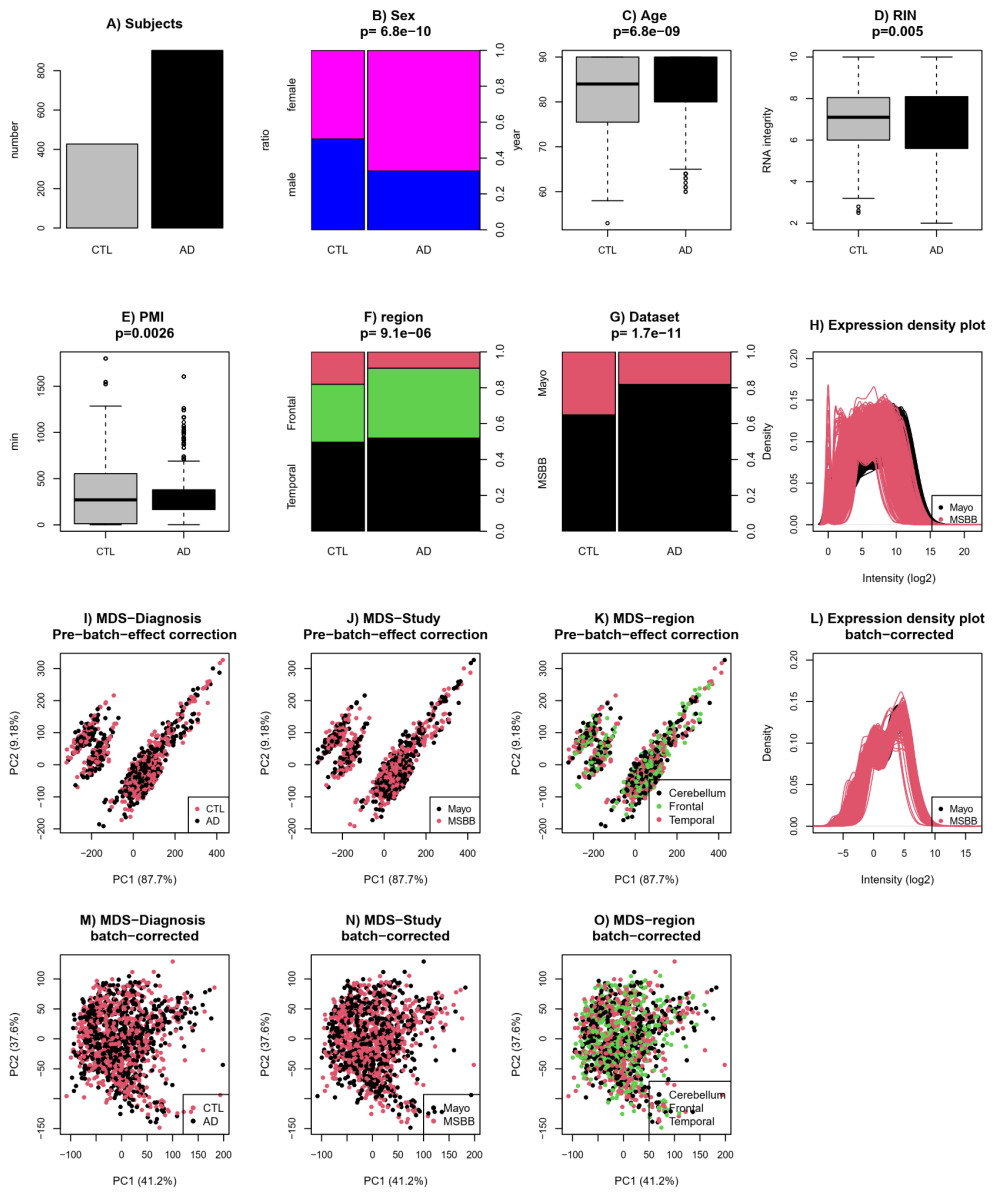

Figure S2. Quality control (QC) plots are shown for AD datasets. Normalized data from the cohorts analyzed in this study consisted of the anterior prefrontal cortex, perirhinal cortex, superior temporal gyrus, pars opercularis, temporal cortex, and cerebellum brain samples from subjects with AD (n = 906) and controls (n = 479). Sample outliers were detected by standardized network connectivity z-scores and removed. Batch effects were corrected for the studies. Multidimensional scaling (MDS) plots show sample clustering by the first two expression principal components. Groups were balanced by available covariates and potential confounding factors.

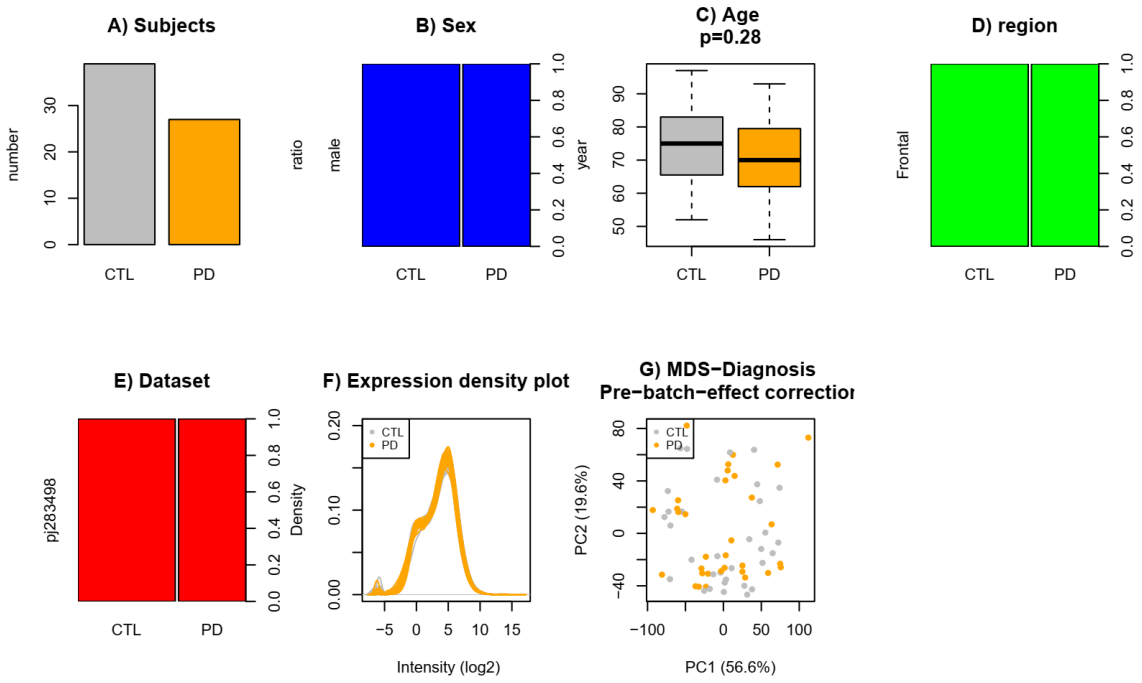

Figure S3. QC plots for the PD dataset. This dataset consisted of dorsolateral prefrontal cortex (BA9; frontal lobe) brain samples from subjects with PD (n = 29) and controls (n = 44). Sample outliers were detected by standardized network connectivity z-scores and removed. Multidimensional scaling (MDS) plots show sample clustering by the first two expression principal components. Groups were balanced by available covariates and potential confounding factors.

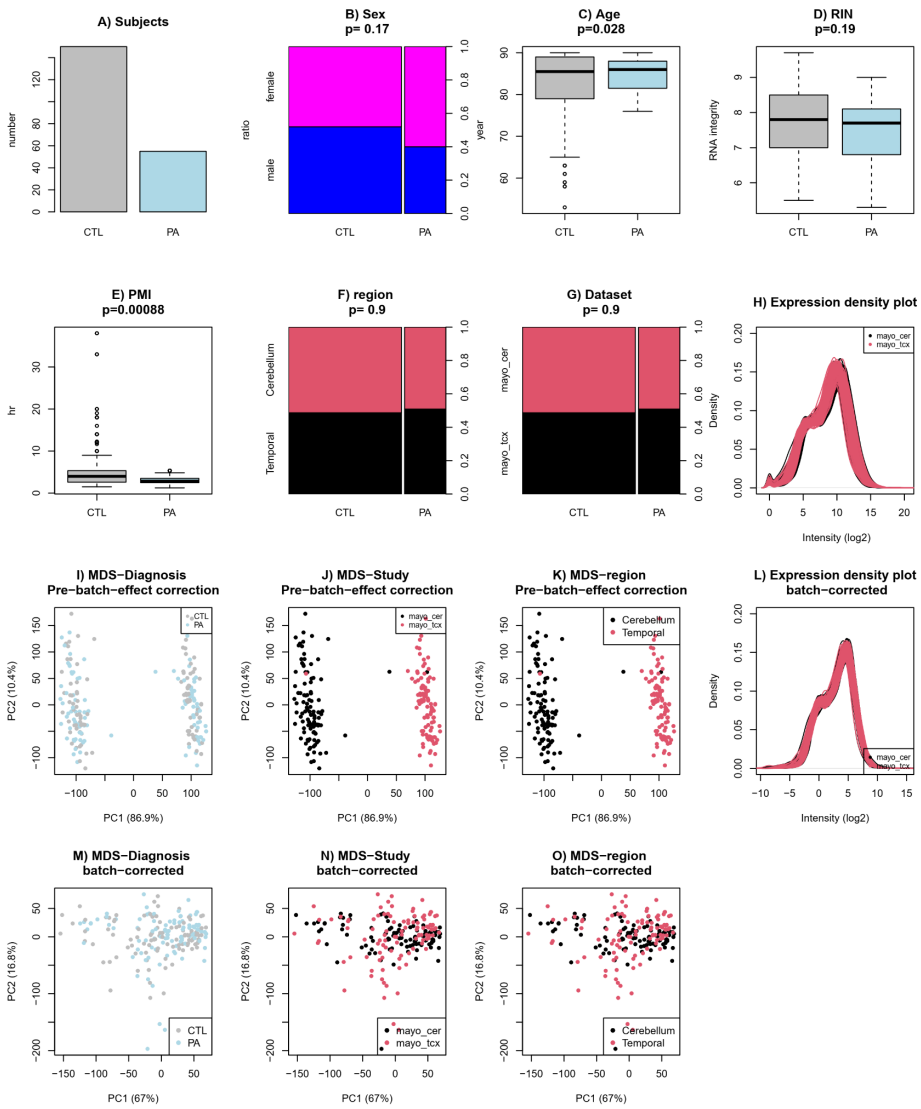

Figure S4. QC plots are shown for PA datasets. Normalized data from one group by two cohorts analyzed here consisted of the temporal cortex and cerebellum brain samples from subjects with PA (n =58) and controls (n =155). Sample outliers were detected by standardized network connectivity z-scores and removed. Batch effects were corrected for the studies. Multidimensional scaling (MDS) plots show sample clustering by the first two expression principal components. Groups were balanced by available covariates and potential confounding factors.

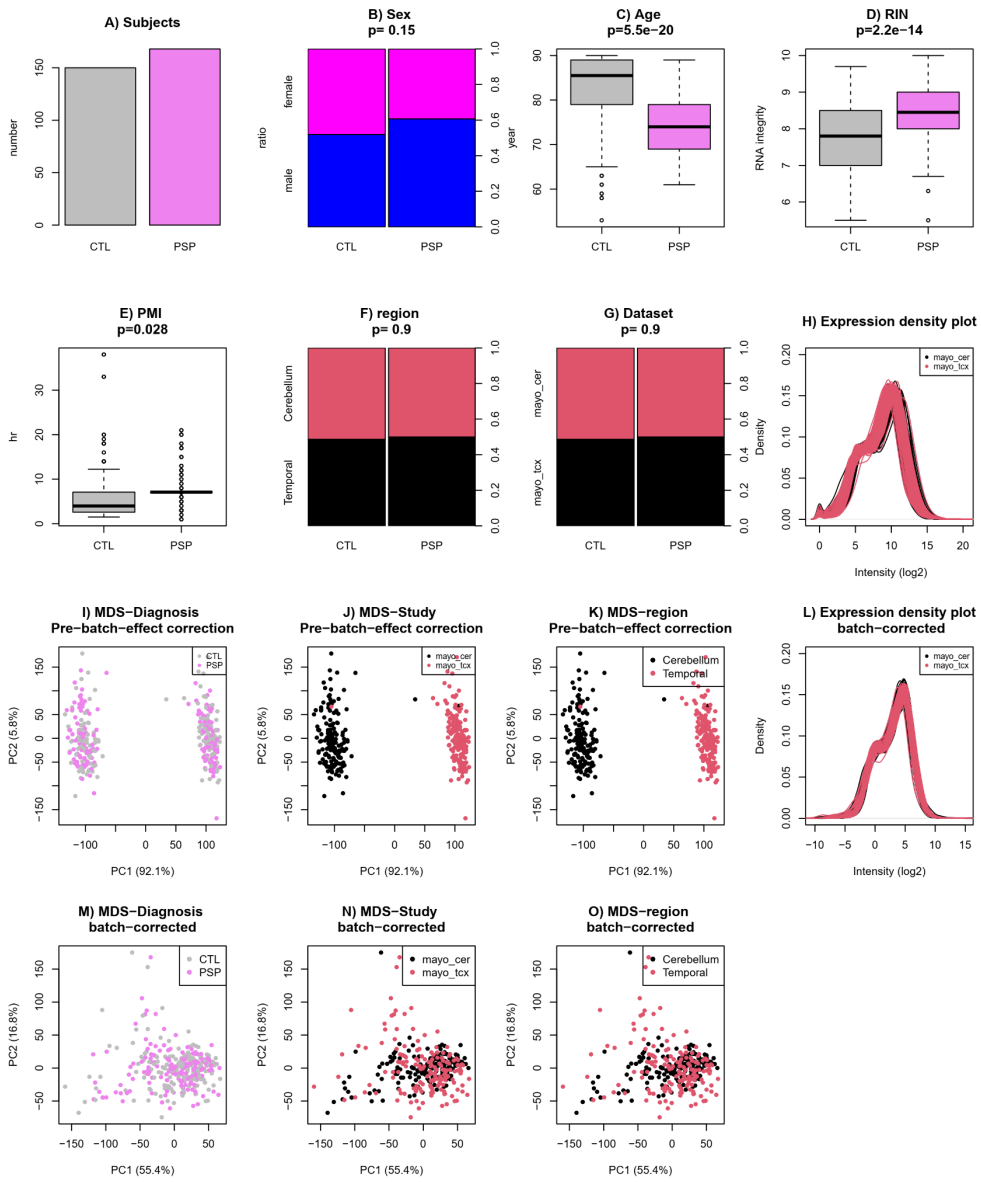

Figure S5. QC plots are shown for PSP datasets. Normalized data from one group by two cohorts analyzed here consisted of the temporal cortex and cerebellum brain samples from subjects with PSP (n = 168) and controls (n =155). Sample outliers were detected by standardized network connectivity z-scores and removed. Batch effects were corrected for the studies. Multidimensional scaling (MDS) plots show sample clustering by the first two expression principal components. Groups were balanced by available covariates and potential confounding factors.

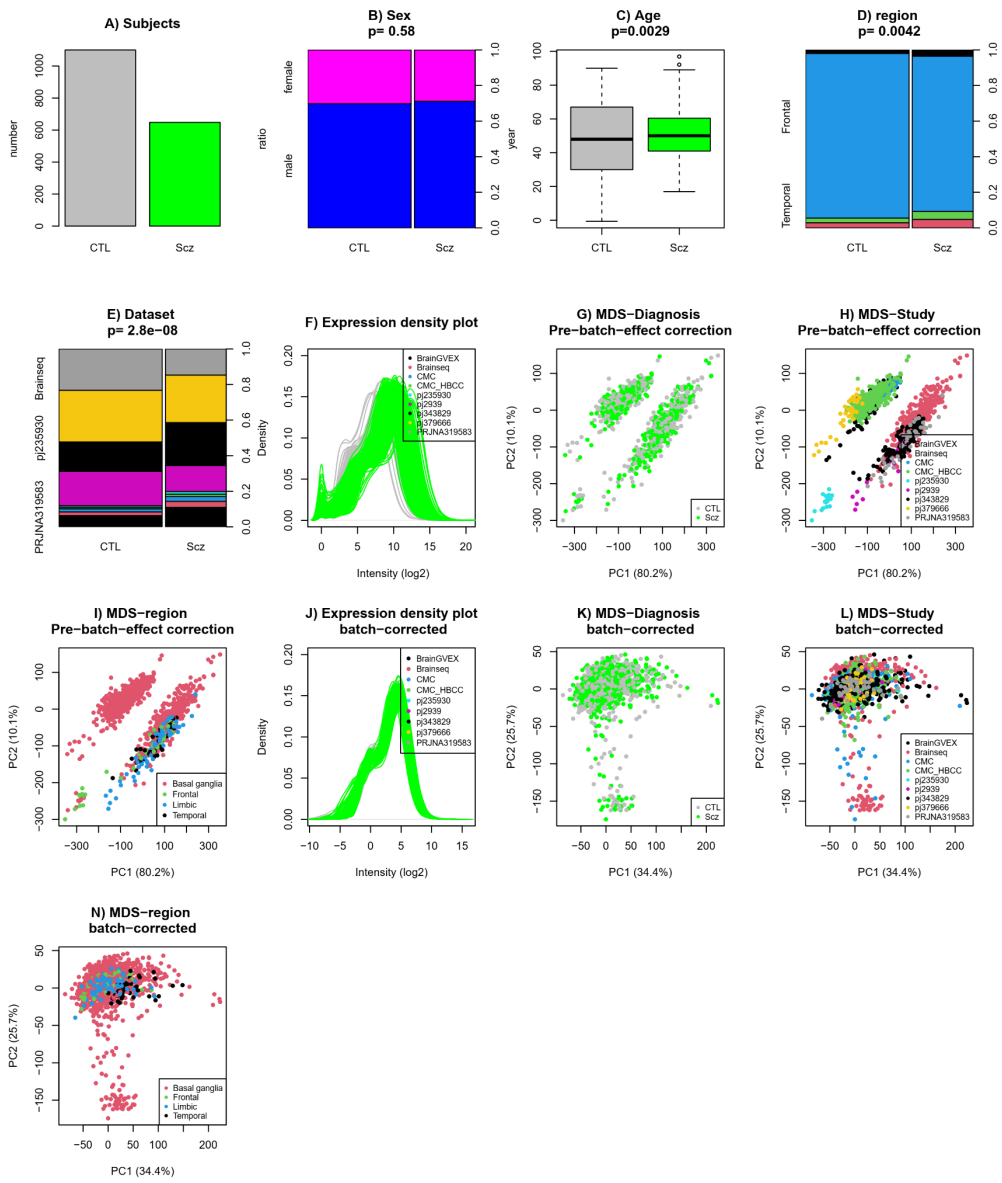

Figure S6. QC plots are shown for Scz datasets. Normalized data from nine studies analyzed here consisted of granular frontal area 9, DLPFC, anterior cingulate cortex, ventral anterior cingulate 25, amygdala, superior temporal gyrus, and nucleus accumbens brain samples from subjects with Scz (n = 535) and controls (n = 1172). Sample outliers were detected by standardized network connectivity z-scores and removed. Batch effects were corrected for the studies. Multidimensional scaling (MDS) plots show sample clustering by the first two expression principal components. Groups were balanced by available covariates and potential confounding factors.

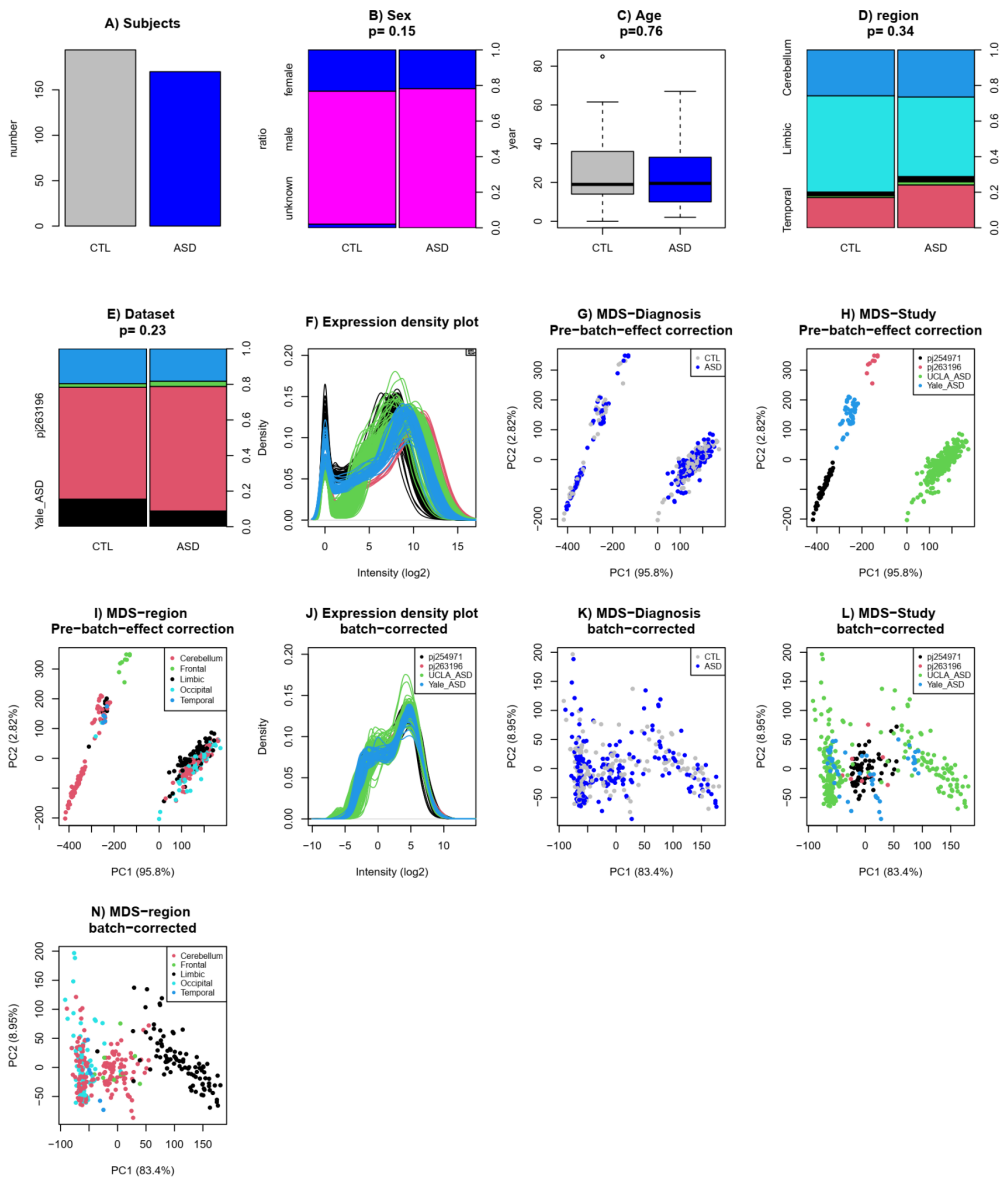

Figure S7. QC plots are presented for ASD datasets. Normalized data from five studies analyzed here consisted of the corpus callosum, inferior temporal cortex, auditory cortex, V1C (primary visual cortex), superior frontal gyrus, and DLPFC brain samples from subjects with ASD (n = 187) and controls (n = 239). Sample outliers were detected by standardized network connectivity z-scores and removed. Batch effects were corrected for the studies. Multidimensional scaling (MDS) plots show sample clustering by the first two expression principal components. Groups were balanced by available covariates and potential confounding factors.

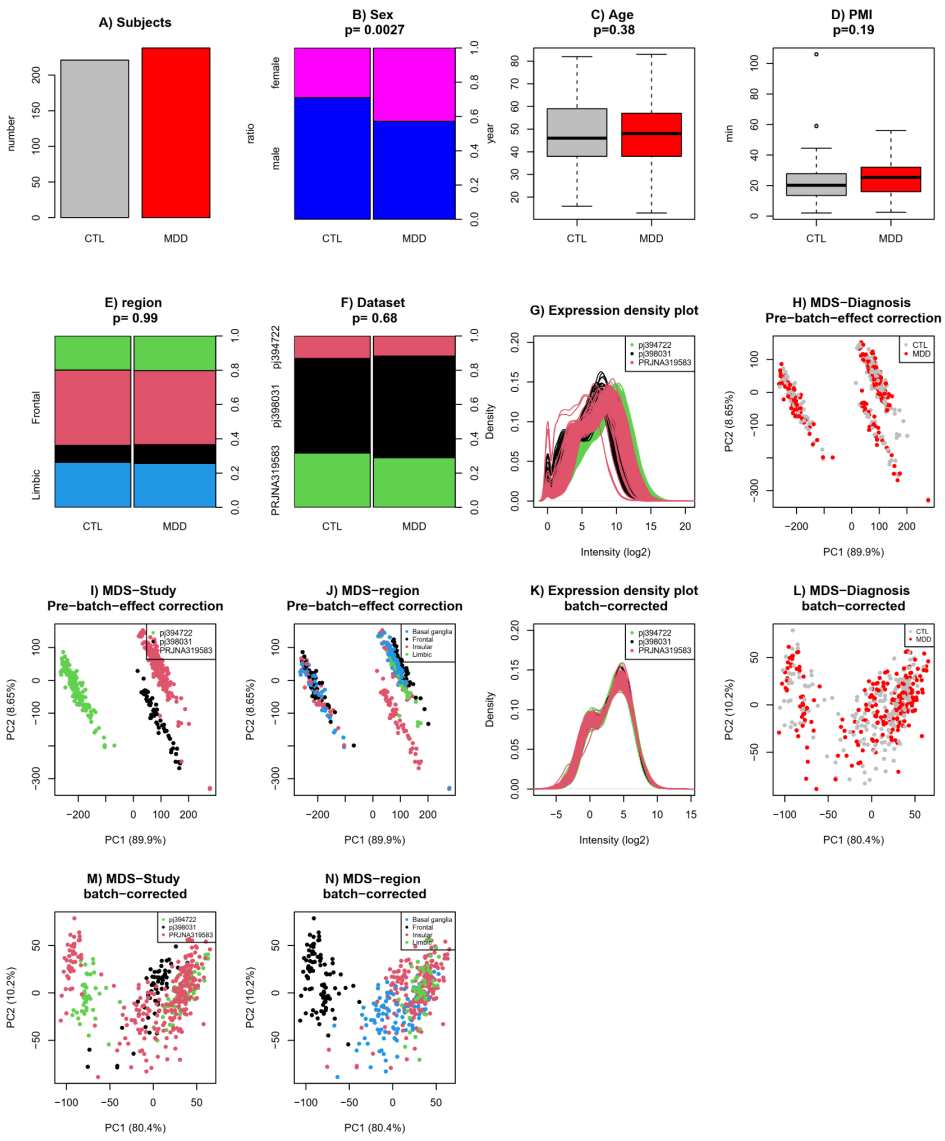

Figure S8. QC plots are shown for MDD datasets. Normalized data from three studies analyzed here consisted of the anterior insula (aINS), nucleus accumbens (Nac), DLPFC, orbitofrontal cortex (OFC), ventral subiculum (vSub), cingulate gyrus 25 (Cg25), anterior cingulate cortex, brain samples from subjects with MDD (n = 240) and controls (n = 221). Sample outliers were detected by standardized network connectivity z-scores and removed. Batch effects were corrected for the studies. Multidimensional scaling (MDS) plots show sample clustering by the first two expression principal components. Groups were balanced by available covariates and potential confounding factors.

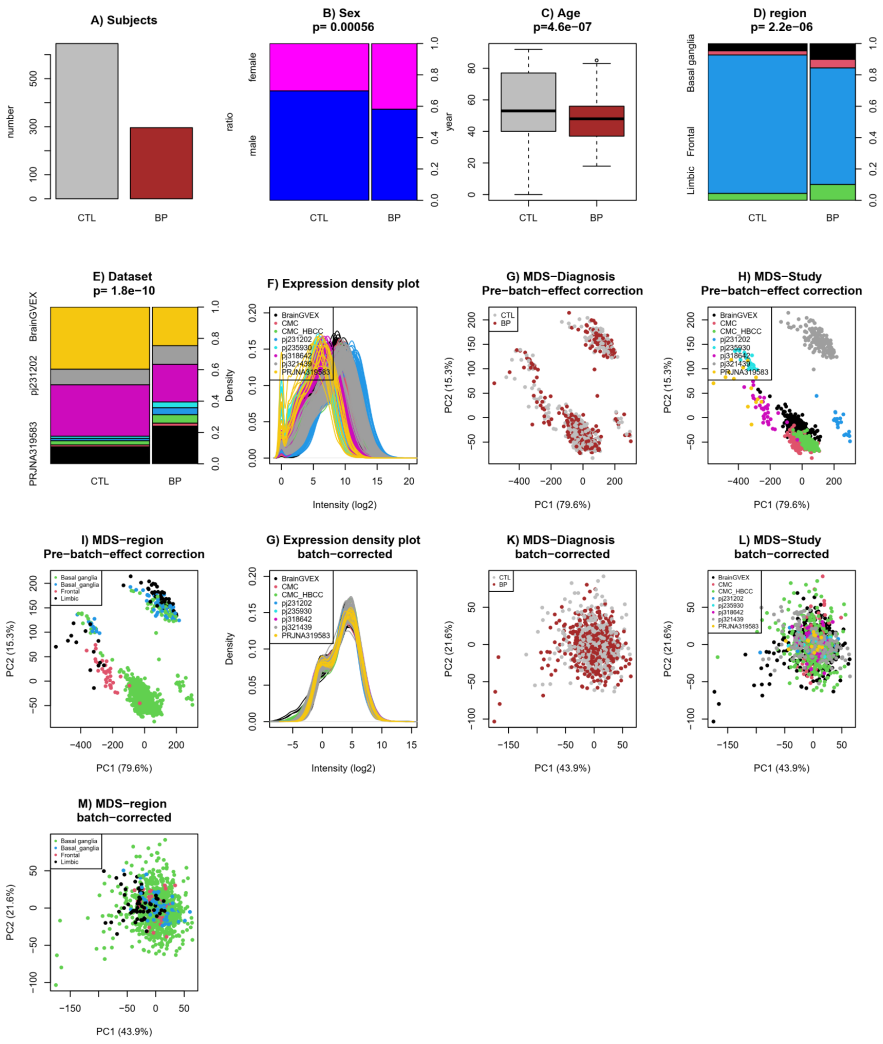

Figure S9. A set of QC plots is shown for BP datasets. Normalized data from eight studies analyzed here consisted of the corpus callosum, inferior temporal cortex, auditory cortex, V1C (primary visual cortex), superior frontal gyrus, and DLPFC brain samples from subjects with BP (n = 511) and controls (n = 837). Sample outliers were detected by standardized network connectivity z-scores and removed. Batch effects were corrected for the studies. Multidimensional scaling (MDS) plots show sample clustering by the first two expression principal components. Groups were balanced by available covariates and potential confounding factors.

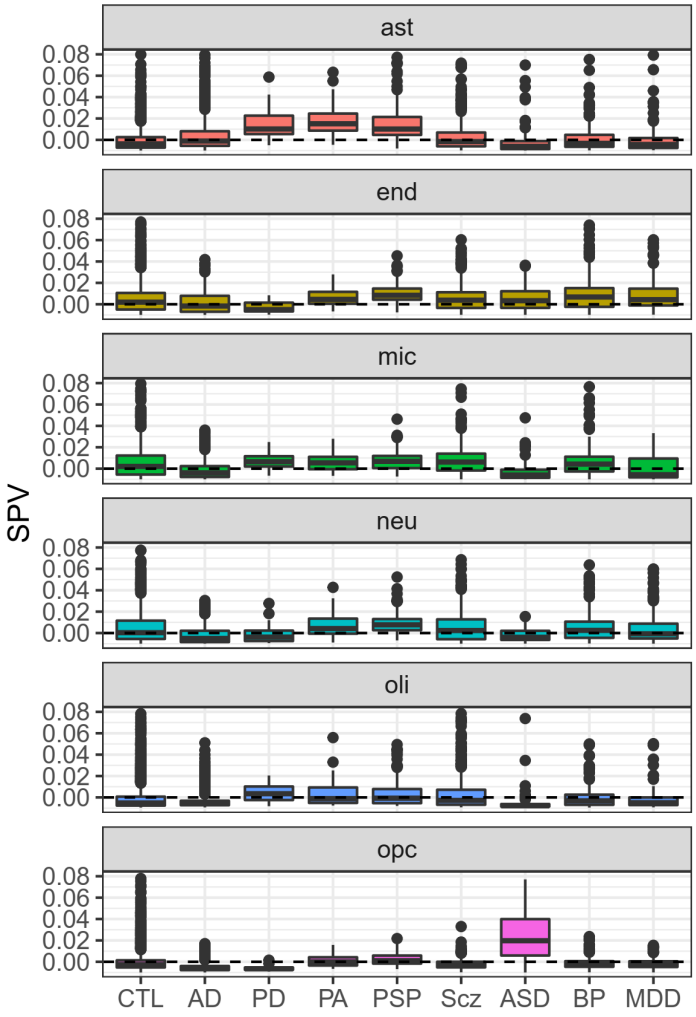

Figure S10. A boxplot showing cell-type proportions calculated for Bulk RNA-Seq data for each condition with the y-axis showing the surrogate cell-type proportion values (SPV). ast; astrocyte, end; endothelial, mic; microglia, neu; neuron, oli; oligodendrocyte, opc; oligodendrocyte progenitor cell.

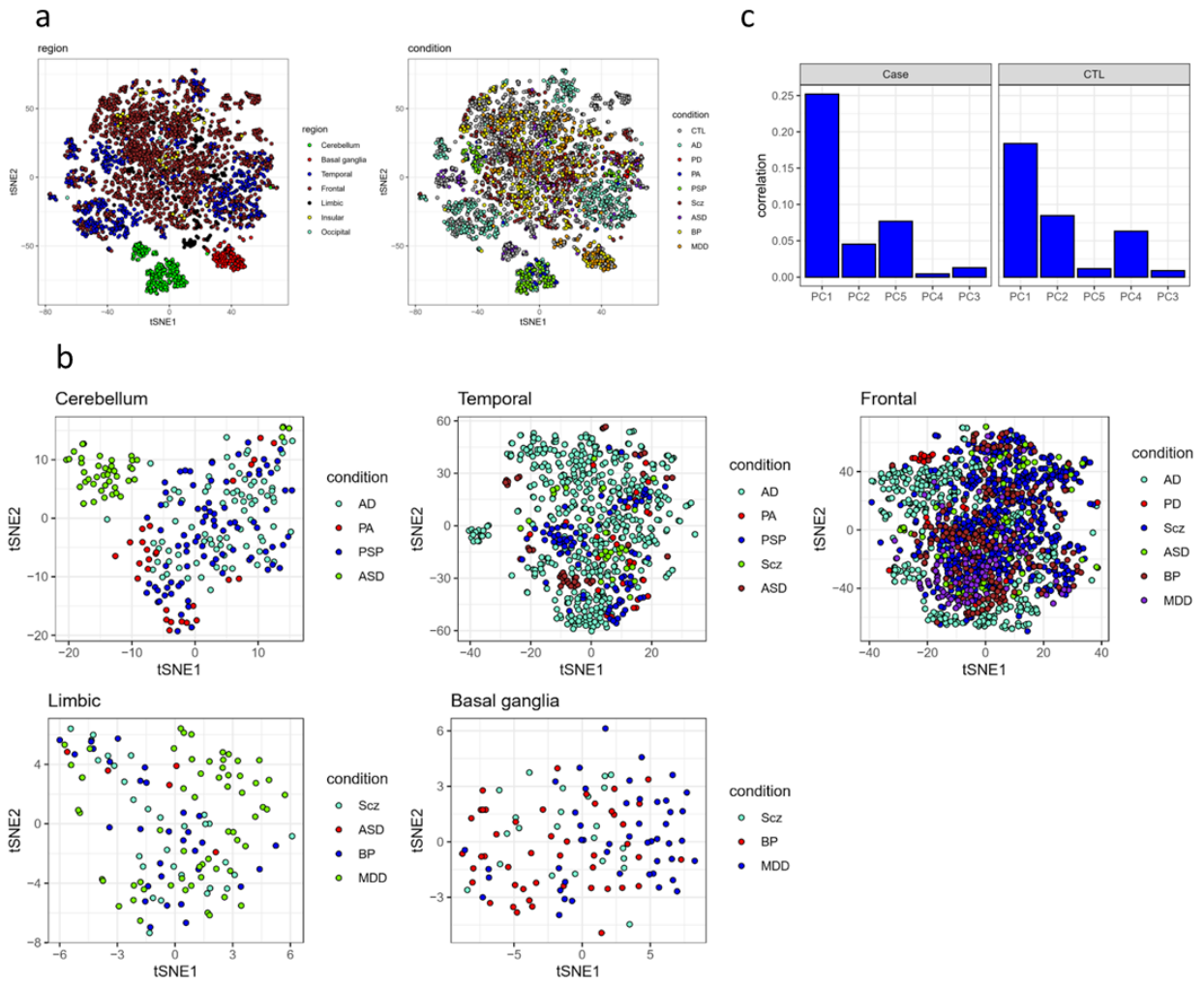

Figure S11. tSNE visualization of the samples used in this study. **a**) tSNE of all pooled samples including controls, colored by region (**left**) and by condition (**right**). **b**) tSNE of the samples from five brain regions for different conditions, excluding controls. **c**) Correlation of top PCs from gene expression with brain region as a covariate for cases and control samples.

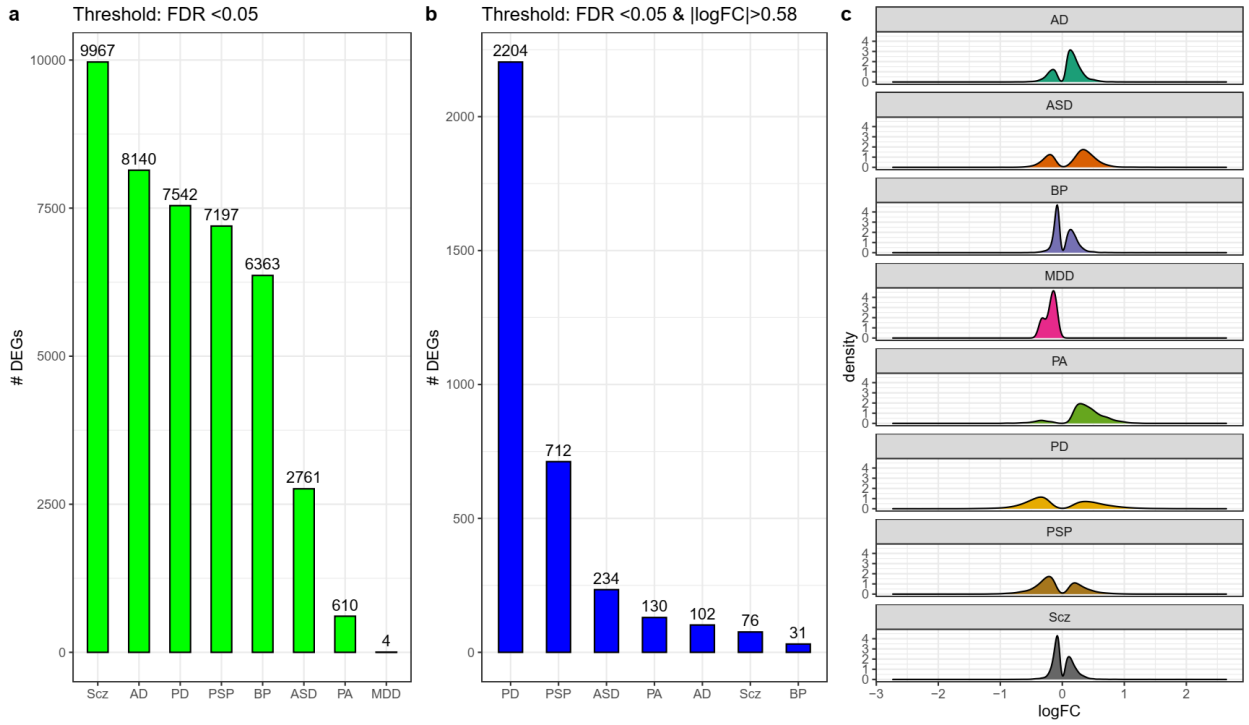

Figure S12. condition-specific transcriptome alterations. The number of differentially expressed genes (DEGs) across conditions with a significant threshold of FDR-corrected P < 0.05 (**a**) and those genes that have an absolute logFC of bigger than 0.58 (**b**). c) A density plot shows the distribution of logFCs across the conditions.

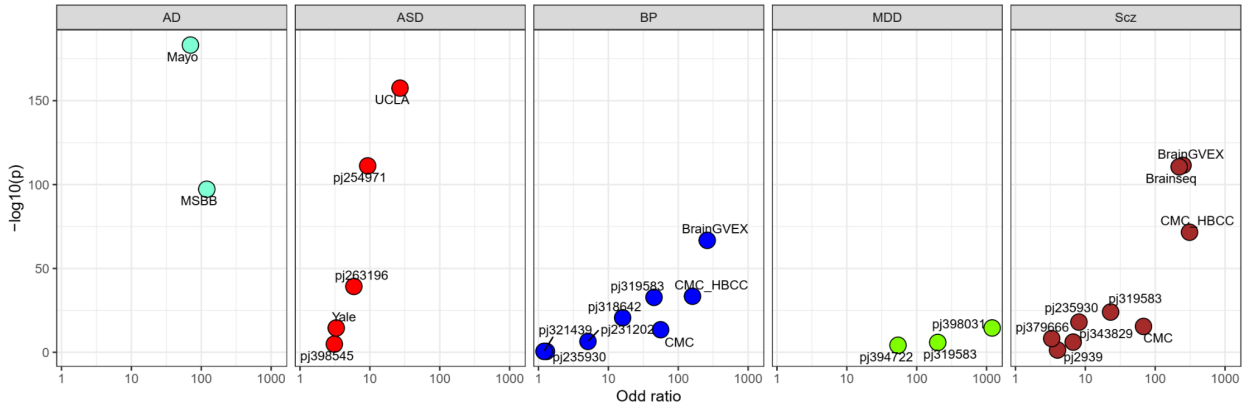

Figure S13. **Reproducibility of differential expression results**. The odd ratio on the *x-axis* shows the similarity ratio of the list of differentially expressed genes for each condition compared to those obtained from individual datasets. (FDR-corrected p-value <0.05)

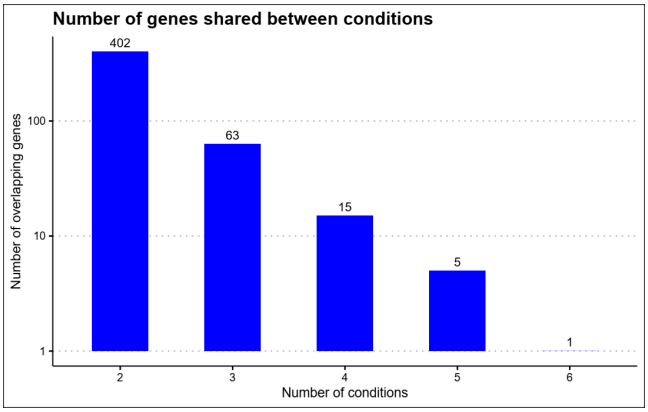

Figure S14. The number of differentially expressed genes shared between at least two conditions.

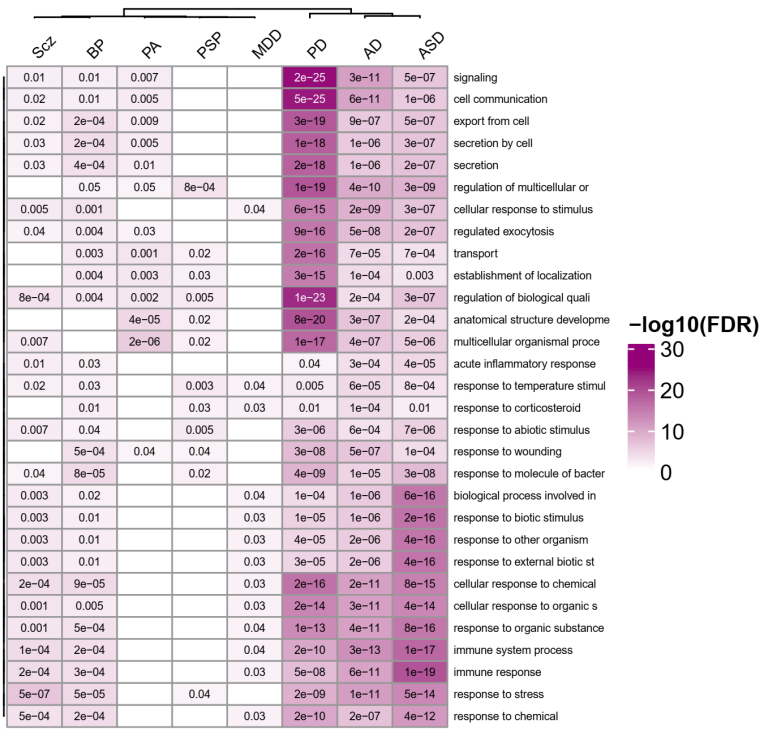

Figure S15. Gene enrichment analysis showed enriched pathways commonly dysregulated across conditions (FDR-corrected p-value < 0.05).

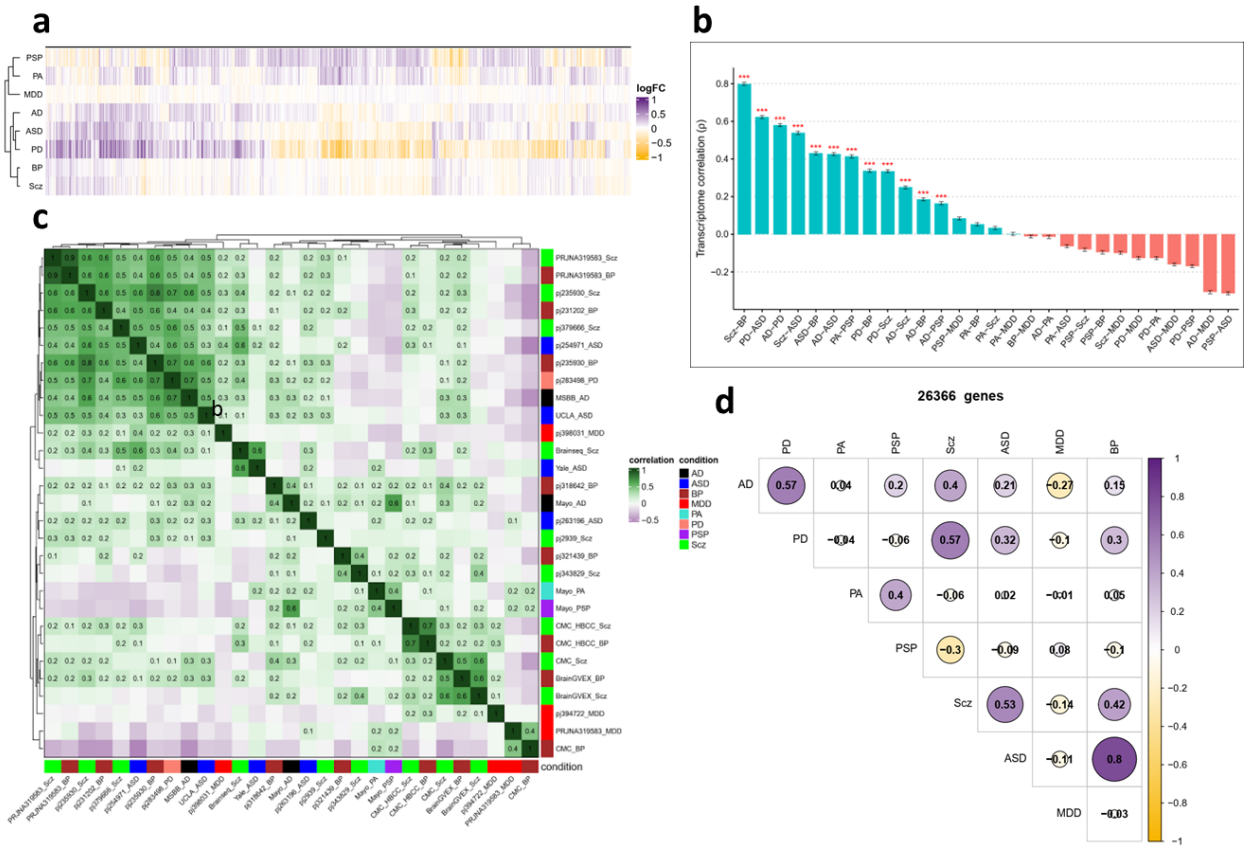

Figure S16. Transcriptome similarities across eight brain diseases. **(a)** A heatmap of logFCs from 15819 genes shared between the conditions. (**b**) Barplot shows top pairwise transcriptome correlation across conditions measured by Spearman's correlation of logFC values from common genes (* FDR-corrected P < 0.05). (**c**) The similarity of transcriptome alterations across eight brain conditions using logFC of 26366 genes including variably expressed genes. (**d**) Reproducibility of transcriptomic correlations. The heatmap shows Spearman's correlations obtained from comparing logFC values of common genes across individual and combined datasets (identified by condition name) for each condition.

	Cerebellum	Temporal	Frontal	Limbic	Occipital	Basal.ganglia	Insular
AD	1911	823	40	–	–	–	–
PD	–	–	1571	–	–	–	–
PA	0	1118	–	–	–	–	–
PSP	2307	911	–	–	–	–	–
Scz	–	0	142	615	–	0	–
ASD	0	147	347	0	0	–	–
BP	–	–	87	1	–	0	–
MDD	–	–	0	0	–	0	0

Figure S17. Number of differentially expressed genes per region for each condition (*FDR-corrected P* < 0.05 & |logFC| > 0.58 was considered as a significant threshold)

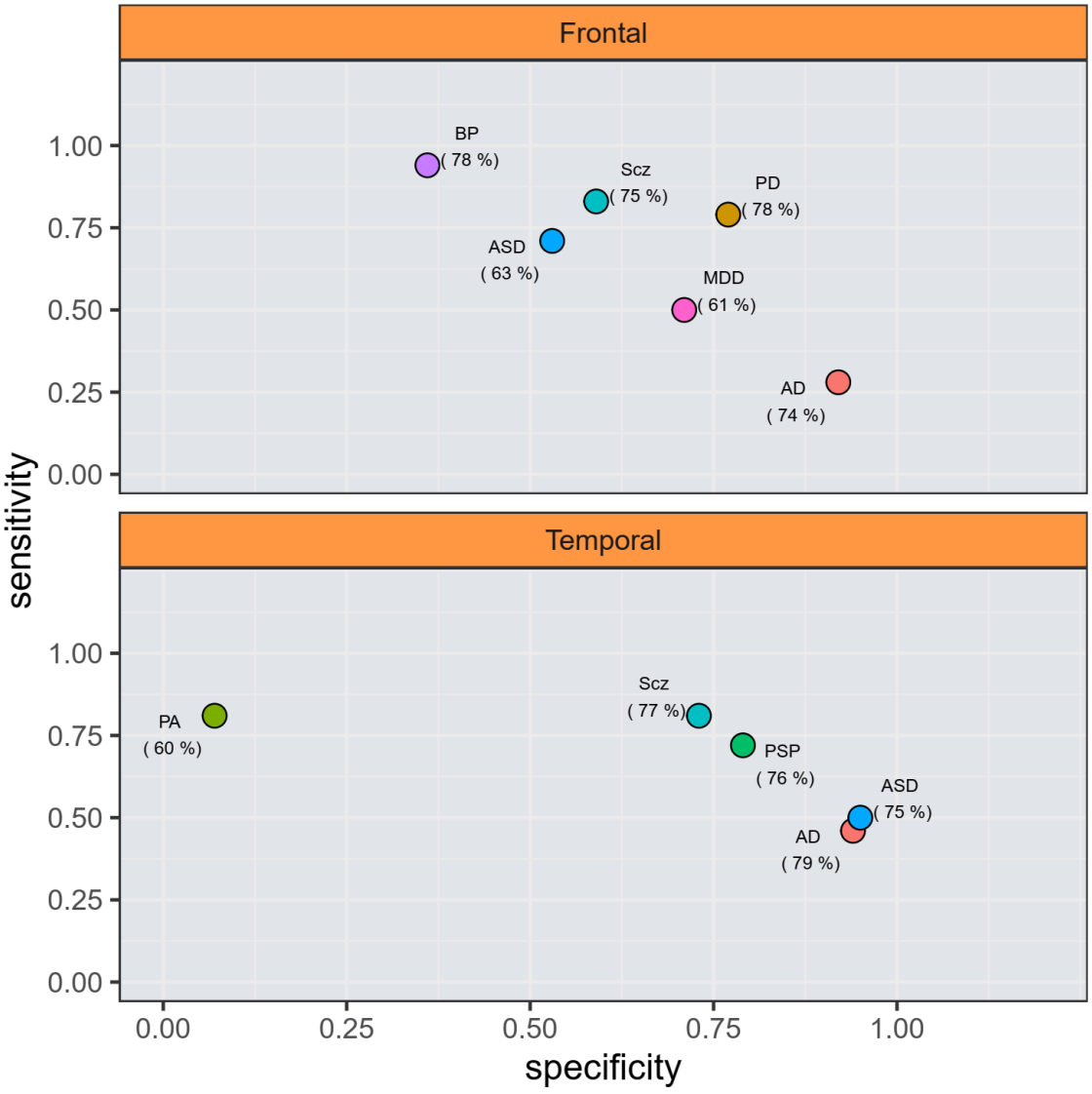

Figure S18. Classifier models show the power of frontal and temporal regions to discriminate between transcriptomes of the conditions and control subjects. Each classifier model was built using the normalized expression of differentially expressed genes for each region. The *x-axis* shows specificity and the *y-axis* shows the sensitivity of the model. Values in parenthesis show prediction accuracy (%) of each region for each condition.

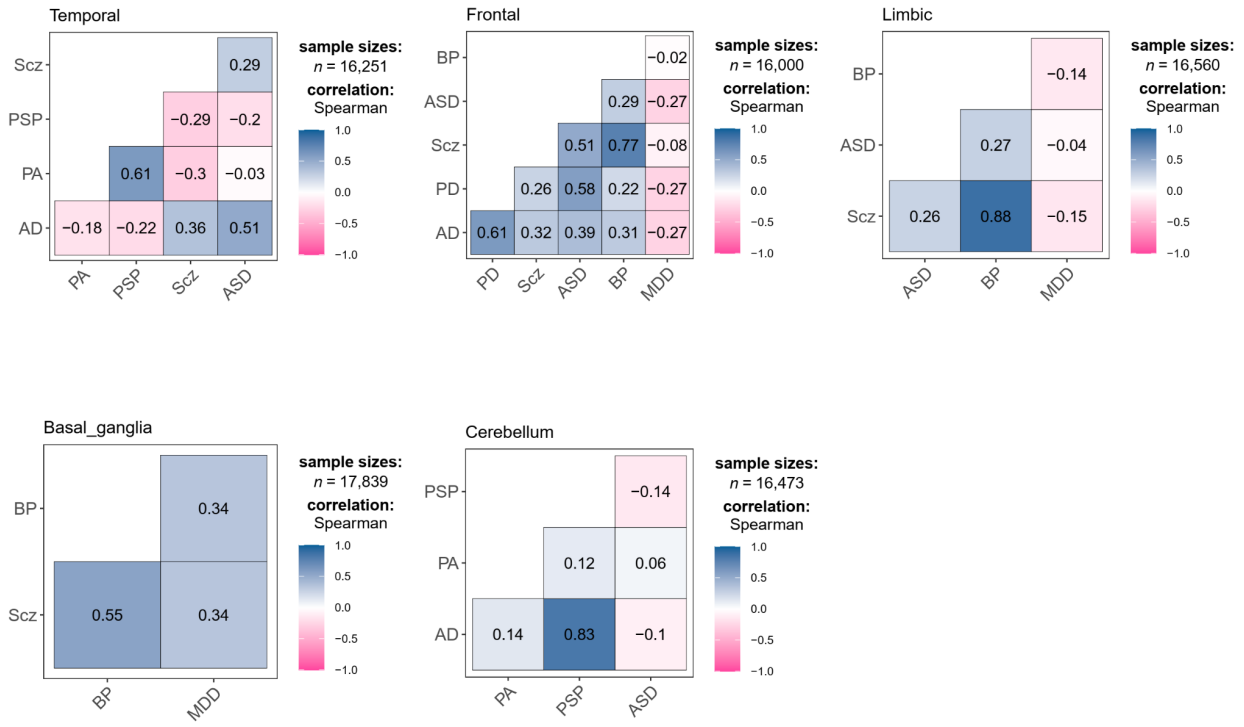

Figure S19. Brain region transcriptome alterations correlations across conditions. Transcriptome alterations overlap within each region obtained by Spearman's correlations using logFC values of the shared genes for each region across conditions.

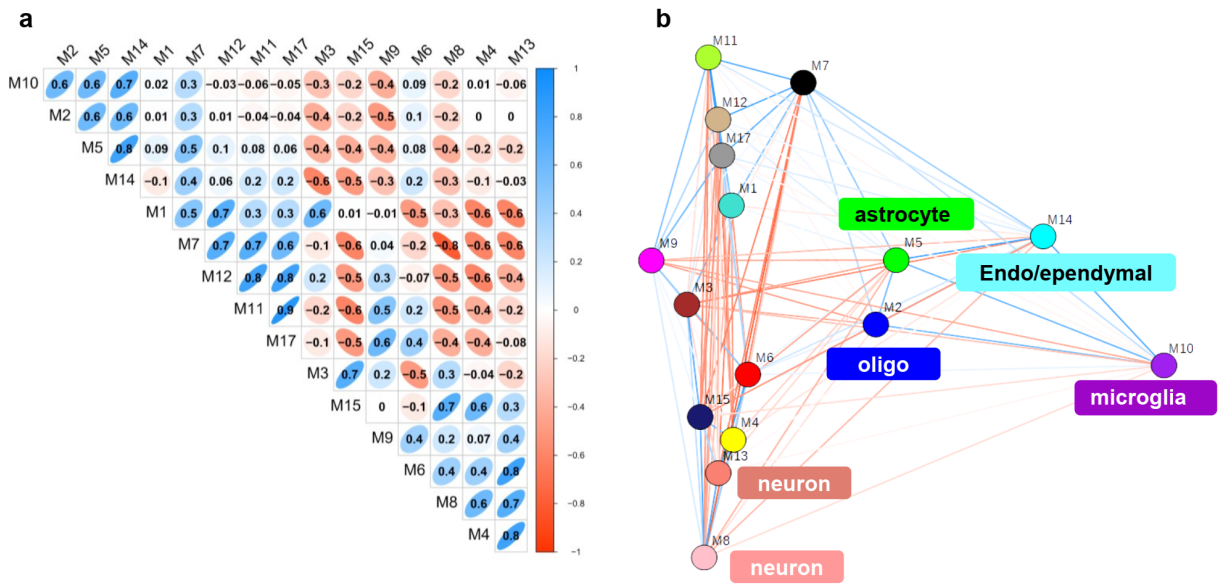

Figure S20. Co-expression modules relationship. (**a**) Correlation plot corresponding to obtained from co-expression topological overlap of the modules. (**b**) A multidimensional scaling plot depicts modules' relationship, with blue and red colors representing positive and negative correlations, respectively.

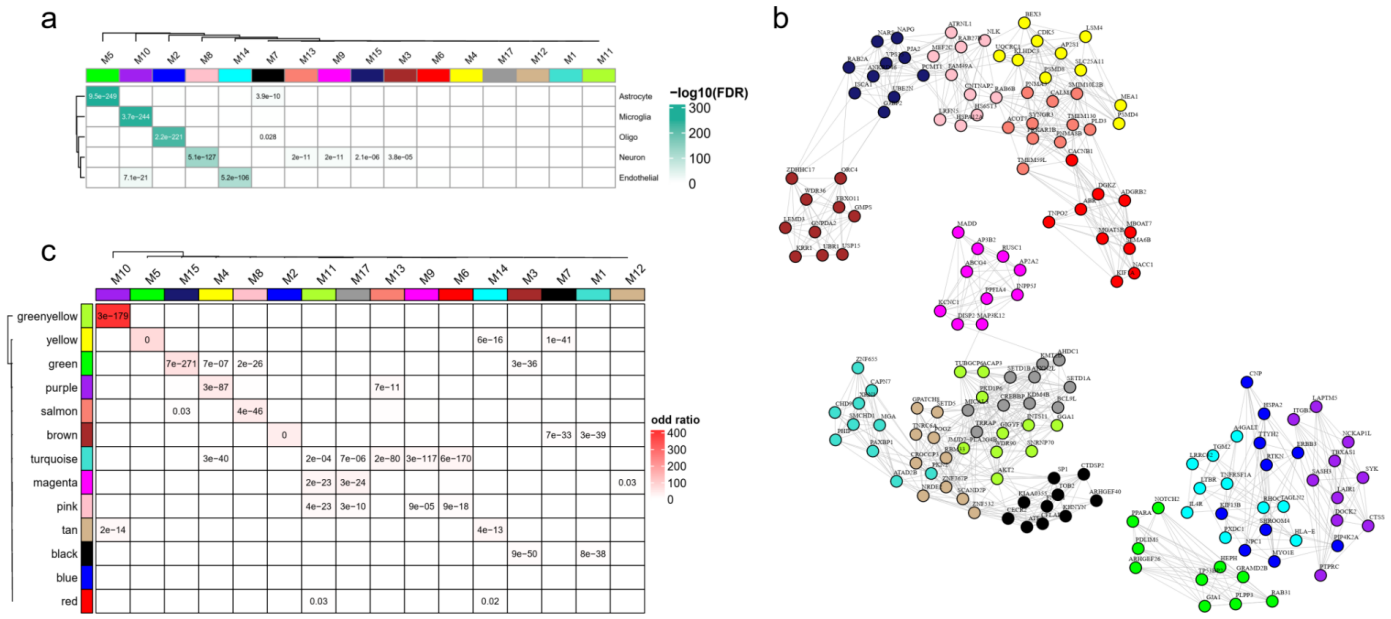

Figure S21. Co-expression modules characteristics. (a) Brain cell type-specific enrichment of co-expression modules was measured using a single-cell expression dataset composed of five main brain cell types including neurons, astrocytes, oligodendrocytes, microglia, and endothelial cells. Fisher's exact tests were used to perform the comparisons. Values show -$\log_{10}$(FDR-corrected p-values). (**b**) A network of top hub genes (nodes) within each module with the highest changes across diseases. Each color shows a module. Edges indicate gene-gene weighted correlations. **(c)** Overlap of gene modules with network modules from Gandal et al[20]. The color key shows an odd ratio and FDR-corrected p-values are shown for significant overlaps (FDR<0.05).

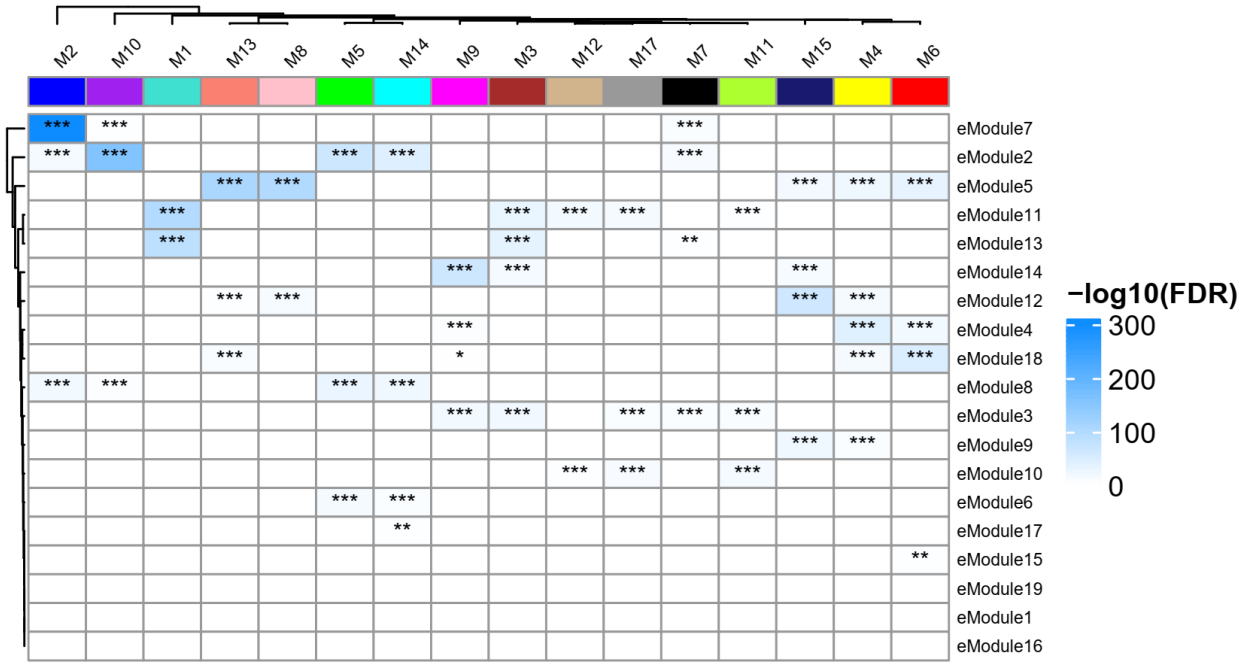

Figure S22. Enrichment of co-expression modules for brain enhancer RNAs.

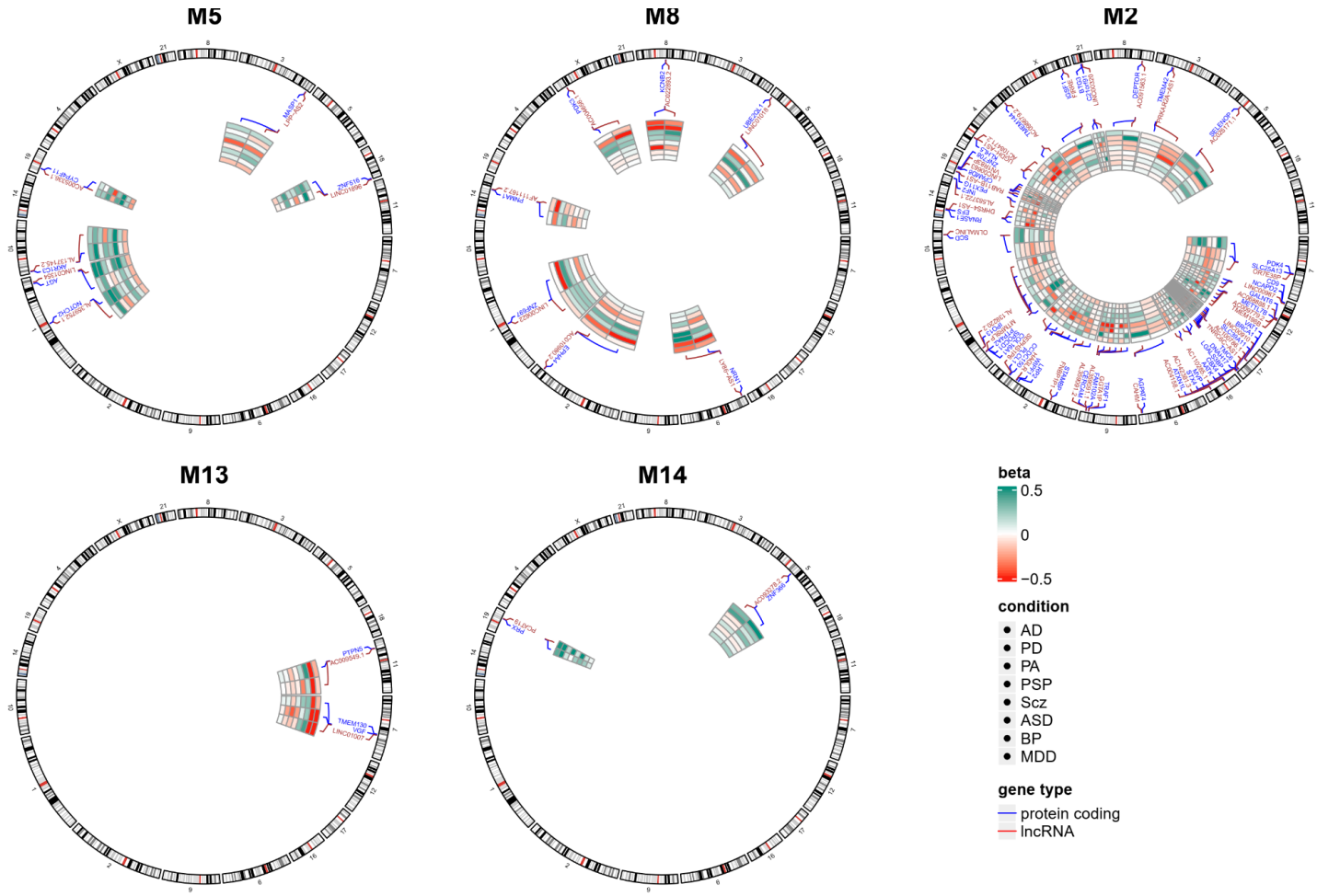

Figure S23. A circular heatmap showing expression fold change (beta) of protein-coding genes and their flanking lncRNAs in cell-type-specific modules across conditions.

References

1. Wang, M. *et al.* The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Scientific Data* vol. 5 (2018).
2. Allen, M. *et al.* Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Scientific Data* vol. 3 (2016).
3. Dumitriu, A. *et al.* Integrative analyses of proteomics and RNA transcriptomics implicate mitochondrial processes, protein folding pathways and GWAS loci in Parkinson disease. *BMC Med. Genomics* **9**, 5 (2016).
4. Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).
5. Xiao, Y. *et al.* The DNA methylome and transcriptome of different brain regions in schizophrenia and bipolar disorder. *PLoS One* **9**, e95875 (2014).
6. Chang, X. *et al.* RNA-seq analysis of amygdala tissue reveals characteristic expression profiles in schizophrenia. *Transl. Psychiatry* **7**, e1203 (2017).
7. Corley, S. M., Tsai, S.-Y., Wilkins, M. R. & Shannon Weickert, C. Transcriptomic Analysis Shows Decreased Cortical Expression of NR4A1, NR4A2 and RXRB in Schizophrenia and Provides Evidence for Nuclear Receptor Dysregulation. *PLoS One* **11**, e0166944 (2016).
8. Wu, J. Q. *et al.* Transcriptome Sequencing Revealed Significant Alteration of Cortical Promoter Usage and Splicing in Schizophrenia. *PLoS ONE* vol. 7 e36351 (2012).
9. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
10. PsychENCODE Consortium *et al.* The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712

(2015).

11. Ramaker, R. C. *et al.* Post-mortem molecular profiling of three psychiatric disorders. *Genome Med.* **9**, 72 (2017).
12. Wright, C. *et al.* Altered expression of histamine signaling genes in autism spectrum disorder. *Transl. Psychiatry* **7**, e1126 (2017).
13. Li, J. *et al.* Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. *Mol. Syst. Biol.* **10**, 774 (2014).
14. Liu, X. *et al.* Disruption of an Evolutionarily Novel Synaptic Expression Pattern in Autism. *PLOS Biology* vol. 14 e1002558 (2016).
15. Pantazatos, S. P. *et al.* Whole-transcriptome brain expression and exon-usage profiling in major depression and suicide: evidence for altered glial, endothelial and ATPase activity. *Mol. Psychiatry* **22**, 760–773 (2017).
16. Labonté, B. *et al.* Sex-specific transcriptional signatures in human depression. *Nat. Med.* **23**, 1102–1111 (2017).
17. Pacifico, R. & Davis, R. L. Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder. *Mol. Psychiatry* **22**, 441–449 (2017).
18. MacMullen, C. M., Fallahi, M. & Davis, R. L. Novel PDE10A transcript diversity in the human striatum: Insights into gene complexity, conservation and regulation. *Gene* **606**, 17–24 (2017).
19. Akula, N. *et al.* RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Mol. Psychiatry* **19**, 1179–1185 (2014).
20. Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).