Supplemental Material

***Title*: Characterizing the nuclear and cytoplasmic transcriptomes in developing and mature human cortex uncovers new insight into psychiatric disease gene regulation**

Authors: Amanda J. Price[1,2], Taeyoung Hwang[1], Ran Tao[1], Emily E. Burke[1], Anandita Rajpurohit[1], Joo Heon Shin[1], Thomas M. Hyde[1,3,4], Joel E. Kleinman[1,3], Andrew E. Jaffe[1,2,3,5,6,7], Daniel R. Weinberger[1,2,3,4,7]

Affiliations:

1. Lieber Institute for Brain Development, Baltimore, MD
2. McKusick Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD
3. Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD
4. Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD
5. Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
6. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
7. Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD

Corresponding Author: Daniel R. Weinberger, 855 N Wolfe St, Ste 300; Baltimore MD 21205.

Phone: 1-410-955-1000; Email: drweinberger@libd.org

Table of Contents

Supplemental Methods

Post-Mortem Brain Samples

Briefly, post-mortem human brain was obtained by autopsy primarily from the Offices of the Chief Medical Examiner of the District of Columbia and the Commonwealth of Virginia, Northern District after informed consent from legal next of kin (protocol 90-M-0142 approved by the NIMH/NIH Institutional Review Board). Brain tissue was stored and dissected at the Clinical Center, NIH, Bethesda, Maryland and at the Lieber Institute for Brain Development in Baltimore, Maryland, and brain material was donated and transferred to the Lieber Institute under an approved Material Transfer Agreement. Post-mortem prenatal brain tissue samples were provided by the National Institute of Child Health and Human Development Brain and Tissue Bank for Developmental Disorders (http://www.BTBank.org/) under contracts NO1-HD-4-3368 and NO1-HD-4-3383. The Institutional Review Board of the University of Maryland at Baltimore and the State of Maryland approved the protocol, and the tissue was donated to LIBD under the terms of a material transfer agreement. Clinical characterization, diagnoses, toxicological analysis, and macro- and microscopic neuropathological examinations were performed on all samples using a standardized protocol approved by the Institutional Review Board of the University of Maryland at Baltimore and the State of Maryland. Subjects with evidence of macro- or microscopic neuropathology, drug use, alcohol abuse or psychiatric illness were excluded.

Defining gene sets associated with brain diseases

Nine brain disease-associated gene sets from a previous publication were used in this study, updated where appropriate to reflect the latest advances (Birnbaum et al. 2014). The reasoning and references are restated here for clarity, and the genes are listed in Table S7.

- Syndromal Neurodevelopmental Disorders (NDD): These genes were identified in a study of 38 patients presenting with syndromal neurodevelopmental disorders as disrupted by chromosomal abnormalities (Talkowski et al. 2012).
- Intellectual disability (ID): These genes were selected from a standardized clinical diagnostic genetic testing panel for intellectual disability (Moeschler et al. 2006).
- Neurodegenerative disorders (Neurodegenerative): This set was comprised of genes implicated in Alzheimer's Disease, Parkinson's Disease, Fronto-temporal Dementia, Amyotrophic Lateral Sclerosis (ALS), Huntington's Disease, and Multiple Sclerosis, including established Mendelian genes and genes from the "Top Results" list for each disease from a systematic meta-analysis of genome-wide association studies, reported as of October 2013 in the following databases: www.alzgene.org, www.pdgene.org, www.msgene.org. This set was updated to include genes listed in Table 1 of a new Genome-wide association study (GWAS) for ALS (Nicolas et al. 2018); candidate genes from tables 1 & 2 of a new GWAS for Parkinson's Disease (Chang et al. 2017); and genes identified in linkage disequilibrium blocks implicated in a new Alzheimer's Disease GWAS (Kunkle et al. 2019).
- Schizophrenia Copy Number Variants (SCZ-CNVs): This set included genes overlapping recurrent CNVs associated with schizophrenia from a series of studies using standardized diagnosis, genome-wide detection, that confirmed structural variations using at least two different methods for discovery and validation, and had sufficient information to permit the identification of duplicate samples (International Schizophrenia Consortium 2008; Stefansson et al. 2008; Walsh et al. 2008; Xu et al. 2008; Kirov et al. 2009, 2012; Levinson et al. 2011; Vacic et al. 2011). Each identified CNV was rare (<1% frequency in the total sample) and autosomal, and met the study's proposed criteria for

"high confidence". Recurrence was defined as identification in multiple cases across studies. CNVs regardless of size were included, 10–100 kb, and >100 kb. The following regions were identified: *1q21.1, 2p16.3, 3q29, 7q36.3, 15q11.2, 15q13.3, 16p11.2, and 22q11.21*.

- Schizophrenia Single Nucleotide Variants (SCZ-SNVs): This set included genes with rare *de novo* SNVs/Indels associated with schizophrenia as identified in exome sequencing studies of probands or affected families (Xu et al. 2008, 2012; Girard et al. 2011; Gulsuner et al. 2013). To be included, the variant must have been missense, nonsense, frameshift, or splice site mutations, and be additionally validated by direct PCR amplification or Sanger sequencing. This gene set was updated from Birnbaum et al. (2014) to include "high evidence" genes from a recent exome study (Genovese et al. 2016).
- Schizophrenia GWAS (SCZ-GWAS): This set included genes contained within linkage disequilibrium blocks associated with SCZ in a new GWAS for this disorder (Pardiñas et al. 2018).
- Bipolar Affective Disorder GWAS (BPAD-GWAS): This set included genes contained within linkage disequilibrium blocks associated with BPAD in a new GWAS for this disorder (Stahl et al. 2019).
- Autism Spectrum Disorder CNVs (ASD-CNVs): This set included genes in recurrent CNVs identified in GWAS for ASD (Sebat et al. 2007; Marshall et al. 2008; Itsara et al. 2010; Pinto et al. 2010; Levy et al. 2011; Sanders et al. 2011). Criteria for publication inclusion was as follows: standardized ASD diagnosis, detection genome-wide, confirmed structural variations, and sufficient information to permit the identification of duplicate samples. Each identified CNV was rare (<1% frequency in the total sample)

and autosomal, and met the study's proposed criteria for "high confidence". Recurrence was defined as identification in multiple cases across studies. CNVs regardless of size were included, 10–100 kb, and >100 kb. The following regions were identified: *1q21.1*, *2p16.3, 3p14.1, 5p15.2, 7q11.23, 7q31.1, 7q36.2, 8p23.3, 9p24.3, 10q11.23–21.1, 12q24.31, 15q11.2–13.1, 15q13.2–13.3, 15q23–24.1, 16p11.2, 16q23.3, 18q22.1, 20q13.33, 22q11.21, 22q13.33*.

- SFARI Gene Database (ASD-SFARI): This set included genes collated in the Simons Foundation Autism Research Initiative Gene Database, accessed October 6, 2019 (Simons Foundation Autism Research Initiative 2019). This set excluded genes labeled as "syndromal" and genes already included in the ASD-CNVs gene set.

Supplemental tables

Table S1: Phenotype and Sequencing Information. Phenotype and sequencing data for the 23 RNA-sequencing samples analyzed in this work. Refer to Tab 2 in Supplemental_Tables.xlsx.

Table S2: Differentially Expressed Genes by Fraction in Adult and Prenatal Samples. Genes that are significantly differentially expressed between nuclear and cytoplasmic fractions in poly(A)+ samples (FDR≦0.05). Refer to Tab 3 in Supplemental_Tables.xlsx.

Table S3: Differentially Expressed Genes by Age in Cytoplasmic and Nuclear Samples. Genes that are significantly differentially expressed between adult and prenatal age groups in poly(A)+ samples (FDR≦0.05). Refer to Tab 4 in Supplemental_Tables.xlsx.

Table S4: Group-Unique A->I editing sites identified in the data. A list of A->I RNA editing sites identified in the 12 poly(A)+ RNA-sequencing samples and their annotation. Refer to Tab 5 in Supplemental_Tables.xlsx.

Table S5: Summary of uniquely edited sites by fraction and age. The number of RNA editing sites found in all samples of a group but not found in a contrasting group (A subset of the RNA editing sites listed in Table S4). For age and fraction individually, the site must be present in all six poly(A)+ samples of a fraction or age group while not in the contrasting six samples. Subset groups of fraction and age must be present in the three samples of a group and not in the contrasting three samples. For example, 159 editing sites were found in all three adult nuclear RNA samples but were not detected in any adult cytoplasmic RNA samples. Refer to Tab 6 in Supplemental_Tables.xlsx.

Table S6: RNA Binding Protein Sequence Motif Enrichment. Results of the RBP motif enrichment analysis where scores are reported for 4 gene sets of cytoplasmic-enriched genes in adult and prenatal, and nuclear genes enriched in adult and prenatal. Refer to Tab 7 in Supplemental_Tables.xlsx.

Table S7: Genes included in each gene set. The list of gene symbols included in the nine neurodegenerative and neurodevelopmental psychiatric disease gene sets defined in the supplemental methods. Refer to Tab 8 in Supplemental_Tables.xlsx.

Table S8: Disease Gene Set Enrichment in Fraction-Regulated Gene Groups (Human Cortical Tissue). Results of the enrichment tests of genes significantly differentially expressed by fraction in the human cortical samples in the nine sets of disease-associated genes. Refer to Tab 9 in Supplemental_Tables.xlsx.

Table S9: Disease Gene Set Enrichment in Fraction-Regulated Gene Groups (ENCODE Cell Lines). Results of the enrichment tests of genes significantly differentially expressed by fraction in the ENCODE samples in the nine sets of disease-associated genes. Refer to Tab 10 in Supplemental_Tables.xlsx.

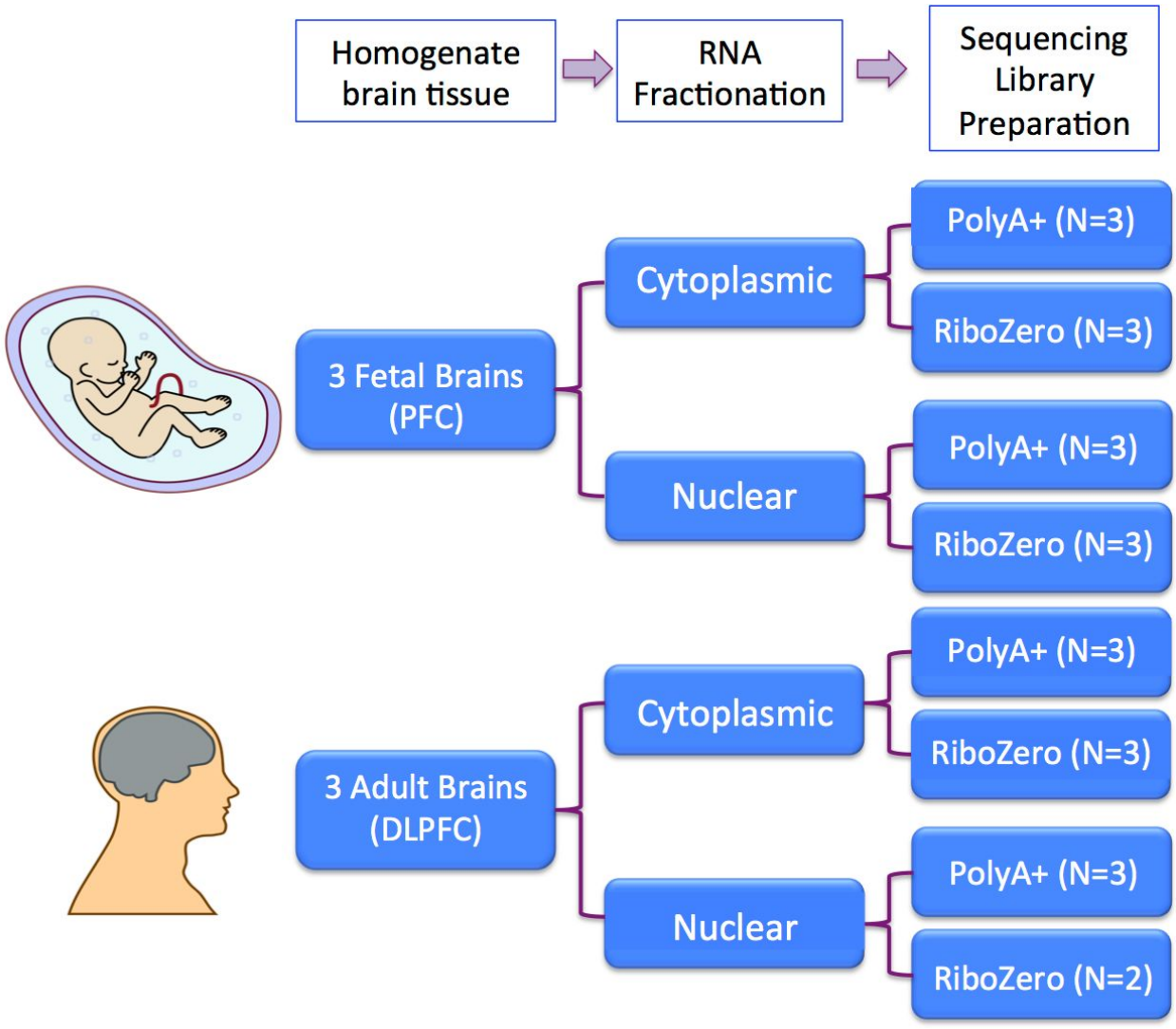

Figure S1: Experimental Design. We characterized the nuclear and cytoplasmic transcriptome in human prenatal postmortem prefrontal cortex (PFC) and adult postmortem dorsolateral prefrontal cortex (DLPFC) using two RNA sequencing library preparation methods. poly(A)+ library preparation selects polyadenylated transcripts via a pull-down step, while Ribo-Zero library preparation relies on a rRNA depletion step.

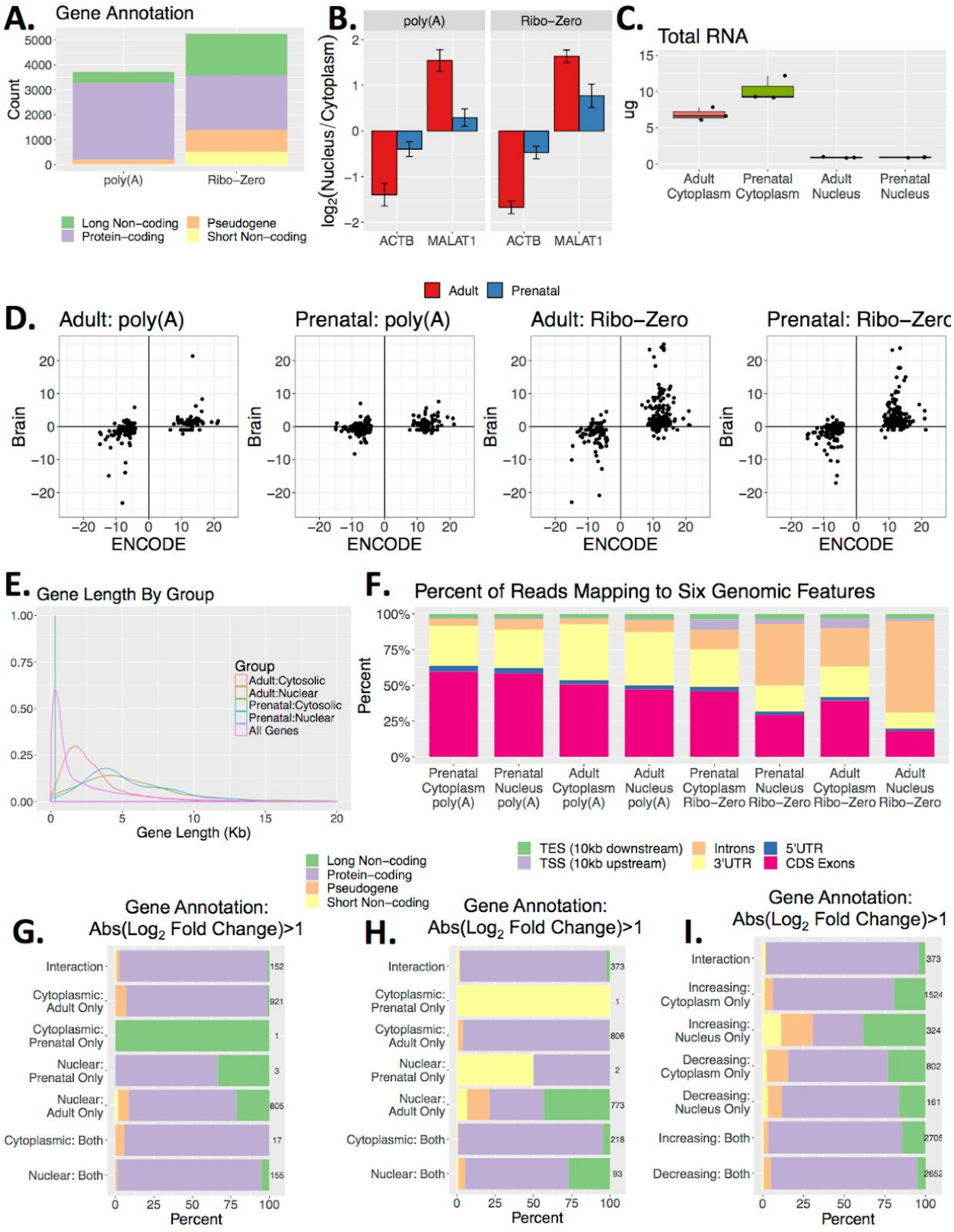
A.
Gene Annotation
Count
poly(A)
Ribo−Zero
Long Non−coding
Protein−coding
Pseudogene
Short Non−coding
B.
poly(A)
Ribo−Zero
log2(Nucleus/Cytoplasm)
ACTB
MALAT1
C.
Total RNA
ug
Adult Cytoplasm
Prenatal Cytoplasm
Adult Nucleus
Prenatal Nucleus
Adult
Prenatal
D.
Adult: poly(A)
Prenatal: poly(A)
Adult: Ribo−Zero
Prenatal: Ribo−Zero
Brain
ENCODE
E.
Gene Length By Group
Group
Adult:Cytosolic
Adult:Nuclear
Prenatal:Cytosolic
Prenatal:Nuclear
All Genes
Gene Length (Kb)
F.
Percent of Reads Mapping to Six Genomic Features
Percent
Prenatal Cytoplasm poly(A)
Prenatal Nucleus poly(A)
Adult Cytoplasm poly(A)
Adult Nucleus poly(A)
Prenatal Cytoplasm Ribo−Zero
Prenatal Nucleus Ribo−Zero
Adult Cytoplasm Ribo−Zero
Adult Nucleus Ribo−Zero
Long Non−coding
Protein−coding
Pseudogene
Short Non−coding
TES (10kb downstream)
TSS (10kb upstream)
Introns
3'UTR
5'UTR
CDS Exons
G.
Gene Annotation:
Abs(Log2 Fold Change)>1
Interaction
Cytoplasmic: Adult Only
Cytoplasmic: Prenatal Only
Nuclear: Prenatal Only
Nuclear: Adult Only
Cytoplasmic: Both
Nuclear: Both
Percent
H.
Gene Annotation:
Abs(Log2 Fold Change)>1
Interaction
Cytoplasmic: Prenatal Only
Cytoplasmic: Adult Only
Nuclear: Prenatal Only
Nuclear: Adult Only
Cytoplasmic: Both
Nuclear: Both
Percent
I.
Gene Annotation:
Abs(Log2 Fold Change)>1
Interaction
Increasing: Cytoplasm Only
Increasing: Nucleus Only
Decreasing: Cytoplasm Only
Decreasing: Nucleus Only
Increasing: Both
Decreasing: Both
Percent

Figure S2: Characterizing the nuclear and cytoplasmic transcriptome in human brain. (A) Differentially expressed genes in nuclear RNA by library type (FDR≤0.05; abs(Log_2 Fold Change)≥1). Ribo-Zero library preparation captures more variation in RNA species in this compartment than poly(A) pull- down. **(B)** *ACTB1*, a known cytoplasmic gene, and *MALAT1*, a known nuclear gene, are enriched in the appropriate subcellular fractions. Error bars reflect standard error. The y-axis shows the log_2 fold change of expression between fractions, stratified by age. **(C)** Micrograms of RNA collected from each sample stratified by age and fraction. A similar amount of RNA being collected from adult and prenatal nuclear samples is evidence that samples underwent appropriate fractionation, despite compartmental differences being more muted in prenatal than adult samples. **(D)** Scatterplots of *t*-statistics for expression differences by fraction as measured in a panel of 296 genes with the greatest expression differences by fraction across ENCODE samples (see methods for how these differences were derived). ENCODE *t*-statistics are on the x-axis, while those measured in homogenate brain stratified by group (left to right: Adult/poly(A), Prenatal/poly(A), Adult/Ribo-Zero, and Prenatal/Ribo-Zero) are on the y-axis. These plots also show that prenatal and adult samples are similarly successfully fractionated. **(E)** Distribution of lengths of genes enriched by fraction in poly(A)+ samples (FDR≤0.05; abs(Log_2 fold change)≥1). “Cytoplasmic” and “Nuclear” reflect the fraction in which the gene is higher expressed, and “Adult” and “Prenatal” represent the age in which the comparison between fractions was made. **(F)** Percent of reads mapping to six genomic features in each group. TES=Transcription end site; TSS=transcription start site; UTR=untranslated region, CDS=coding; kb=kilobase. **(G-H)** Annotation of groups of genes differentially expressed **(G)** by fraction in poly(A), **(H)** by fraction in Ribo-Zero, and **(I)** by age in Robo-Zero (FDR≤0.05; abs(Log_2 fold change)≥1). The total number of genes in each group is listed to the right of each bar.

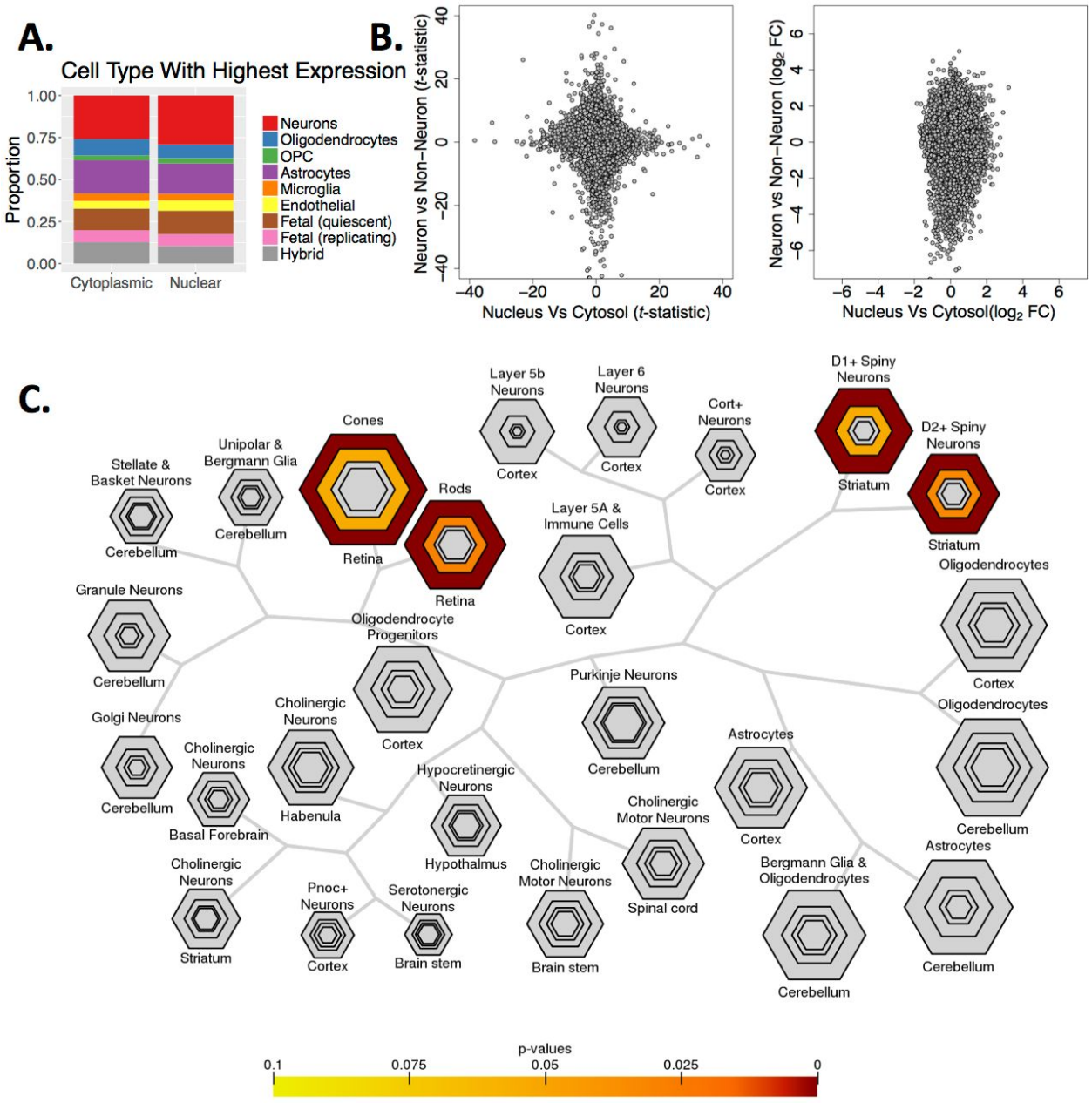

Figure S3: Cell type enrichment in fraction-regulated genes. (A) The proportion of cytoplasmic- and nuclear-enriched genes (FDR≤0.05) identified in the 23 homogenate fractionated samples that are maximally expressed in each cell type as measured in single cell RNA-seq data from human cortex (Darmanis et al. 2015). We assigned each cytoplasmic and nuclear gene a cell type based on which type had the highest expression as measured in the

single cell data. Cell types are as assigned in the single cell data as in the previous publication. **(B)** Scatterplots of *t*-statistics (left) and $\log_2$ fold change (right) as measured between fractions in homogenate cortex in this study (x-axis) and between neurons and non-neurons (ie, NeuN+ and NeuN- nuclear RNA-seq; y-axis). The biggest distinction between brain cell types is that of neurons and non-neurons (primarily glia). Using NeuN, a neuronal marker, one can label neuronal nuclei isolated from postmortem brain tissue and collect neuron- and non-neuron-enriched populations from tissue. The proportion of neurons to non-neurons is dynamic across brain development, which may confound the fractioned gene expression results. Here we show that there is no discernable relationship between neuron/non-neuron expression differences and the nuclear/cytoplasmic differences we identify in homogenate samples, suggesting that these results are not broadly confounded by developmental shifts in cellular composition. **(C)** Results of cell-specific enrichment analysis (CSEA) for the set of genes preferentially expressed in adult nucleus (ie, “Nuclear in Adult Only” genes). Retinal rod and cone cells and striatal DRD1+ and DRD2+ medium spiny neurons are enriched at a specificity index threshold of 0.05.

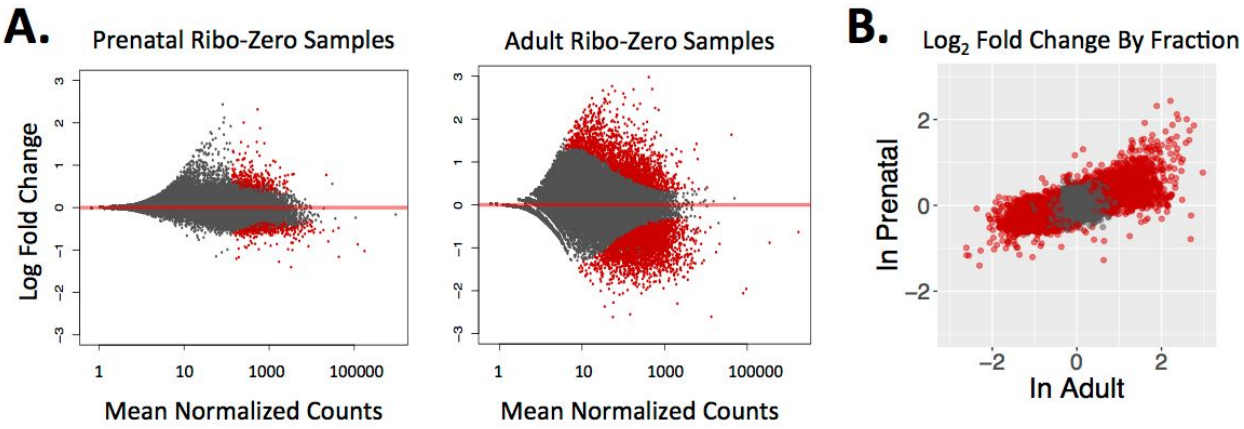

Figure S4: Comparing Fraction and Age in Ribo-Zero Samples. (A) MA plots of prenatal and adult gene expression differences measured across fraction. Red dots indicate FDR≤0.05. **(B)** Log_2 fold change of expression across fraction in adult samples plotted against prenatal samples. Red dots indicate FDR≤0.05.

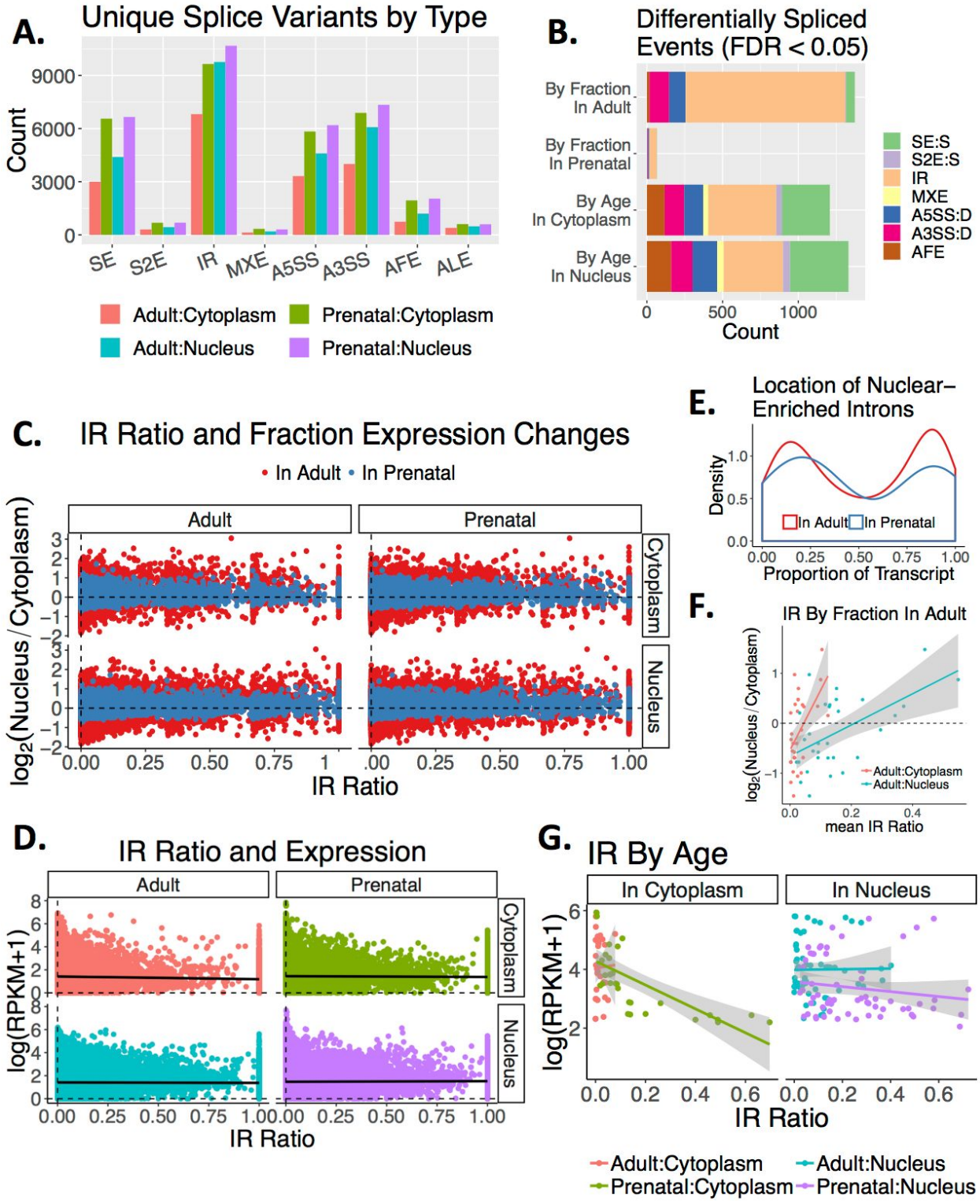

Figure S5: Alternative splicing patterns and intron retention. (A) Counts of unique skipped exons (SE), skipping of two exons (S2E), intron retention (IR), mutually exclusive exons (MXE),

alternative 5' exon splice site use (A5SS), alternative 3' exon splice site use (A3SS), alternative first exon use (AFE) and alternative last exon use (ALE) stratified by fraction and age. Skipped exons (SE) and intron retention (IR) represented the greatest percentage of unique splice variants identified (50.8%). 72.9% more unique splice variants were identified in prenatal than adult samples, suggesting that the higher proportion of spliced RNA products in prenatal samples shown in Fig. 3A is accompanied by greater splicing diversity. By fraction, 42.8% more unique splice variants were identified in nuclear than cytoplasmic RNA. **(B)** Number of differentially expressed splicing events in different comparisons, stratified by splice variant type. As in overall gene expression, prenatal fractions showed fewer significantly differentially expressed splice variants than between adult fractions. Whether a splice variant was more expressed in nuclear than cytoplasmic RNA related to its variant type: significantly differentially expressed IR events by fraction (FDR≤0.05) were more likely to be more abundant in the nucleus (OR=50.9, FDR=8.7×10^{-96}), while SE, distal alternative 5' exon start site (A5SS.D) and 3' exon start site (A3SS.D) events were less abundant in the nucleus (OR=0.091, FDR<1.7×10^{-7}). **(C)** The $\log_2$ fold change of gene expression between fractions in adult (red) and prenatal (blue) plotted against the mean maximum IR ratio for an intron within each gene for each age:fraction group. Samples are stratified by panels indicating the age and fraction of the mean maximum IR ratio. 152,432 introns were shared between all 12 poly(A)+ samples. Across samples, 58.68-85.33% of filtered introns were constitutively spliced in each sample, and 12.20-34.63% had an IR ratio (i.e., intronic reads divided by total intron and flanking exon reads) of greater than zero but less than five percent. **(D)** Gene expression measured as the log of the reads per kilobase per million mapped (RPKM) plus one plotted against the maximum IR ratio for each gene within each sample. Samples are colored and stratified into panels indicating their age:fraction group. The black line depicts the linear regression for each group. **(E)** Density

plot of the location in a transcript of introns higher expressed in the nuclear compartment in adult and prenatal brain, listed as the proportion of the transcript length between the intron and the end of the transcript (3'<5'). Using the Audic and Claverie test for differential intron expression we found 35 introns significantly differentially retained within a transcript by fraction in adults, six by fraction in prenatal cortex, 10 by age in cytoplasmic RNA, and 21 by age in nuclear RNA (FDR≤0.05). Of the fraction-regulated introns, all were more retained (less spliced) in the nuclear compartment. These differentially retained introns tended to be single rather than clustered within a gene (90.4%) and were significantly shorter than the pool of total introns tested ($t<-32.8$, $FDR<2.2\times10^{-25}$). Most relevant to this figure, fraction-regulated introns clustered at the beginning and end of the gene. **(F)** The $\log_2$ fold change of gene expression between fractions as measured in adult plotted against the mean IR ratio for introns differentially retained by fraction in adult (FDR≤0.05) in adult cytoplasmic samples (red) and adult nuclear samples (blue). The linear regression with standard error is shown for both groups. **(G)** The IR ratio of introns differentially retained by age when measured in cytoplasmic and nuclear samples, plotted by gene expression measured as the log of the RPKM plus one. Samples are colored by their age:fraction group. The linear regression line for each group is shown with shading to indicate the standard error.

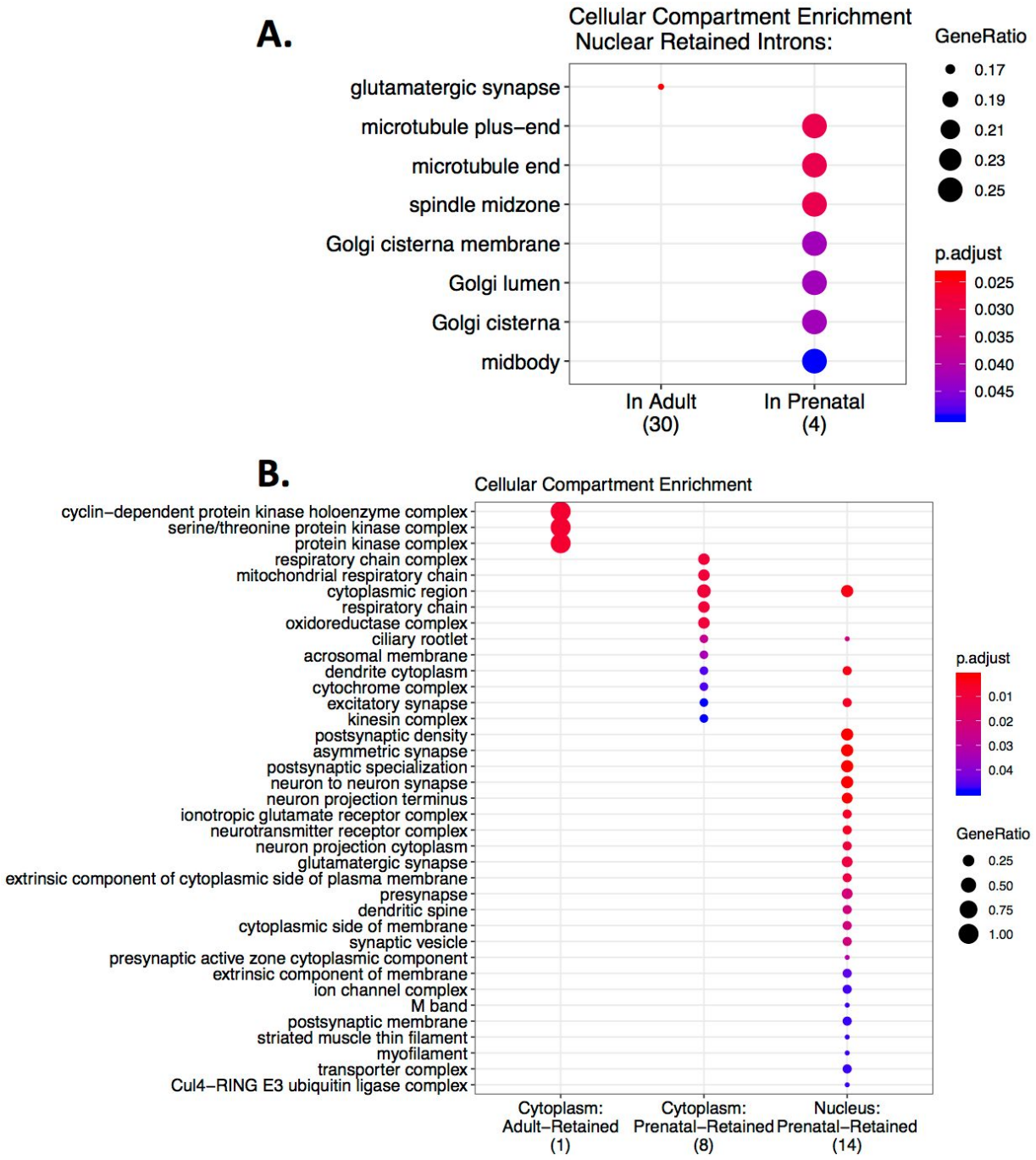

Figure S6: Cellular compartment ontology enrichment for retained introns. (A) Cellular compartment ontology terms enriched in genes containing introns that were higher expressed in nuclear than cytoplasmic RNA in adult (left) and prenatal (right) samples. **(B)** Terms enriched in genes containing introns more retained in adult than prenatal (far left) or more in prenatal than

adult (middle and right columns) when measured in cytoplasm (left and middle columns) or nuclear RNA (right column).

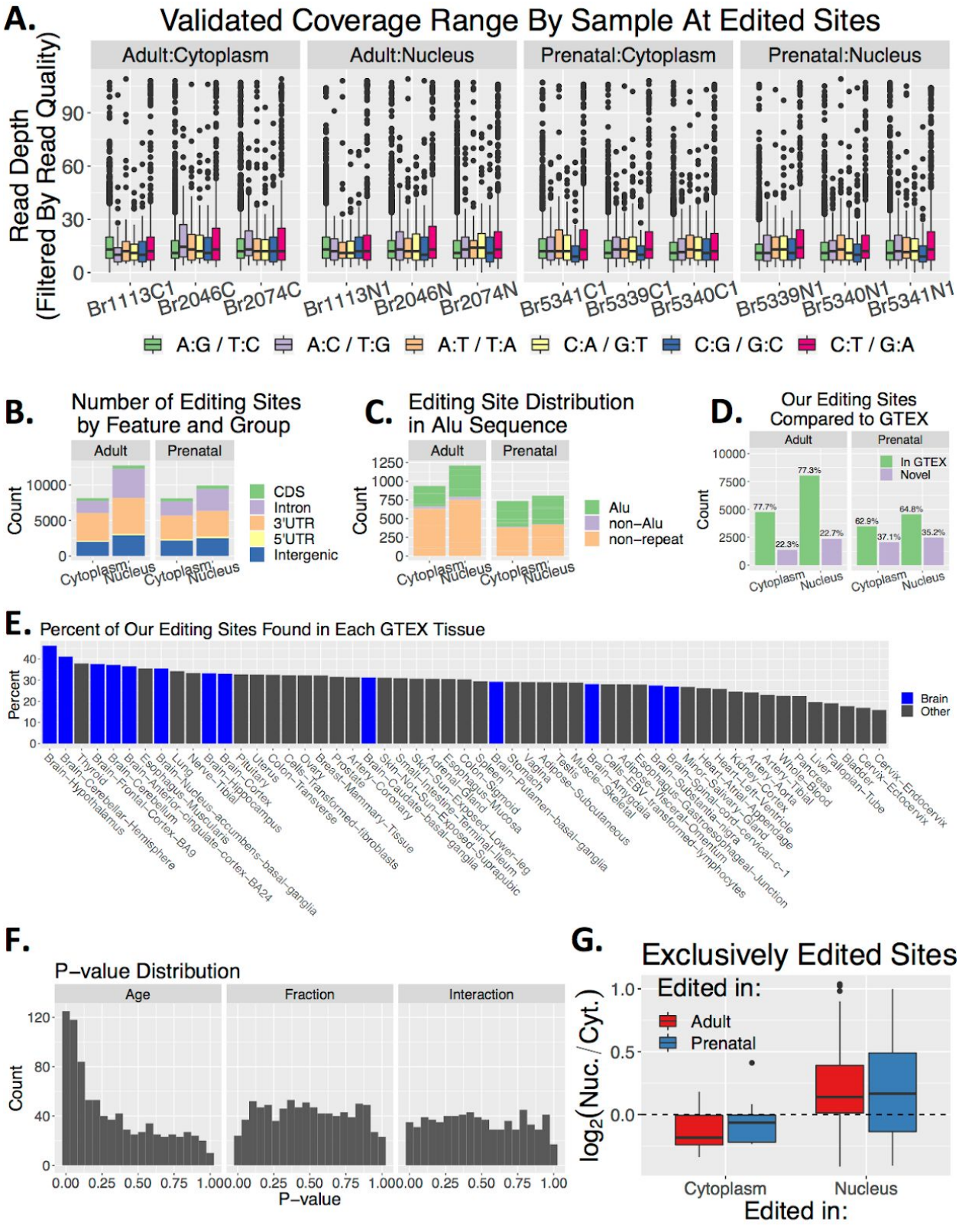
A.
Validated Coverage Range By Sample At Edited Sites
Adult:Cytoplasm
Adult:Nucleus
Prenatal:Cytoplasm
Prenatal:Nucleus
Read Depth (Filtered By Read Quality)
Br1113C1
Br2046C
Br2074C
Br1113N1
Br2046N
Br2074N
Br5341C1
Br5339C1
Br5340C1
Br5339N1
Br5340N1
Br5341N1
A:G / T:C
A:C / T:G
A:T / T:A
C:A / G:T
C:G / G:C
C:T / G:A
B.
Number of Editing Sites by Feature and Group
Adult
Prenatal
Count
Cytoplasm
Nucleus
CDS
Intron
3'UTR
5'UTR
Intergenic
C.
Editing Site Distribution in Alu Sequence
Alu
non−Alu
non−repeat
D.
Our Editing Sites Compared to GTEX
In GTEX
Novel
77.7%
22.3%
77.3%
22.7%
62.9%
37.1%
64.8%
35.2%
E.
Percent of Our Editing Sites Found in Each GTEX Tissue
Percent
Brain
Other
F.
P−value Distribution
Age
Fraction
Interaction
P−value
G.
Exclusively Edited Sites
Edited in:
Adult
Prenatal
log2(Nuc. / Cyt.)
Cytoplasm
Nucleus

Figure S7: Highly edited gene products contain RNA editing sites unique to an age/fraction group that associate with expression levels. (A) Distribution of filtered read depth per sample for each editing context. “A:G/T:C” sites are considered A-to-I editing sites in our unstranded RNA-sequencing data. Read coverage was fairly even over all samples at edited sites, with a median coverage of 11-12 reads per site across samples. **(B)** Annotation of RNA features that contain an A-to-I editing site stratified by fraction and age. In line with previous reports, 21.7-33.8% fell within intronic sequence and 37.6-50.8% within 3’UTR sequence in each fraction and age tested. **(C)** Number of A-to-I editing sites that overlap an Alu sequence, a non-Alu repeat, or no repeats. 40.0-42.0% of A-to-I editing sites overlapped an Alu repeat sequence. **(D)** Number and percentage of A-to-I editing sites identified in our data that are also identified in GTEX consortium data. Of the previously unreported editing sites, 43.1% more were detected in nuclear than cytoplasmic RNA, and 13.8% more were detected in prenatal than adult samples. **(E)** Percentage of editing sites identified in each GTEX tissue that are also identified in our data. 69% of our 18,907 A-to-I editing sites were also detected in Genotype-Tissue Expression (GTEx) project data (Tan et al. 2017), particularly in GTEx brain samples (46.3%). **(F)** Distribution of unadjusted p-values calculated from linear regression assessing editing rate changes by age adjusting for fraction, fraction adjusting for age, or age:fraction interaction effects in the 1,025 sites found in all samples. After adjusting for false discovery rate, 81 sites were associated with age, while only 9 were associated with fraction and 6 with an interaction between age and fraction. **(G)** Log_2 fold change of expression by fraction as measured in prenatal samples, in genes that include an editing site present in all cytoplasmic but no nuclear samples (right) or all nuclear and no cytoplasmic samples (left) in either adult (red) or prenatal (blue). The 159 adult nuclear editing sites were more likely to fall

within intronic sequence than cytoplasmic sites (OR=5.9, FDR=9.8×10^{-3}), which were more likely to fall within 3'UTR sequence (OR=3.96, FDR=9.8×10^{-3}).

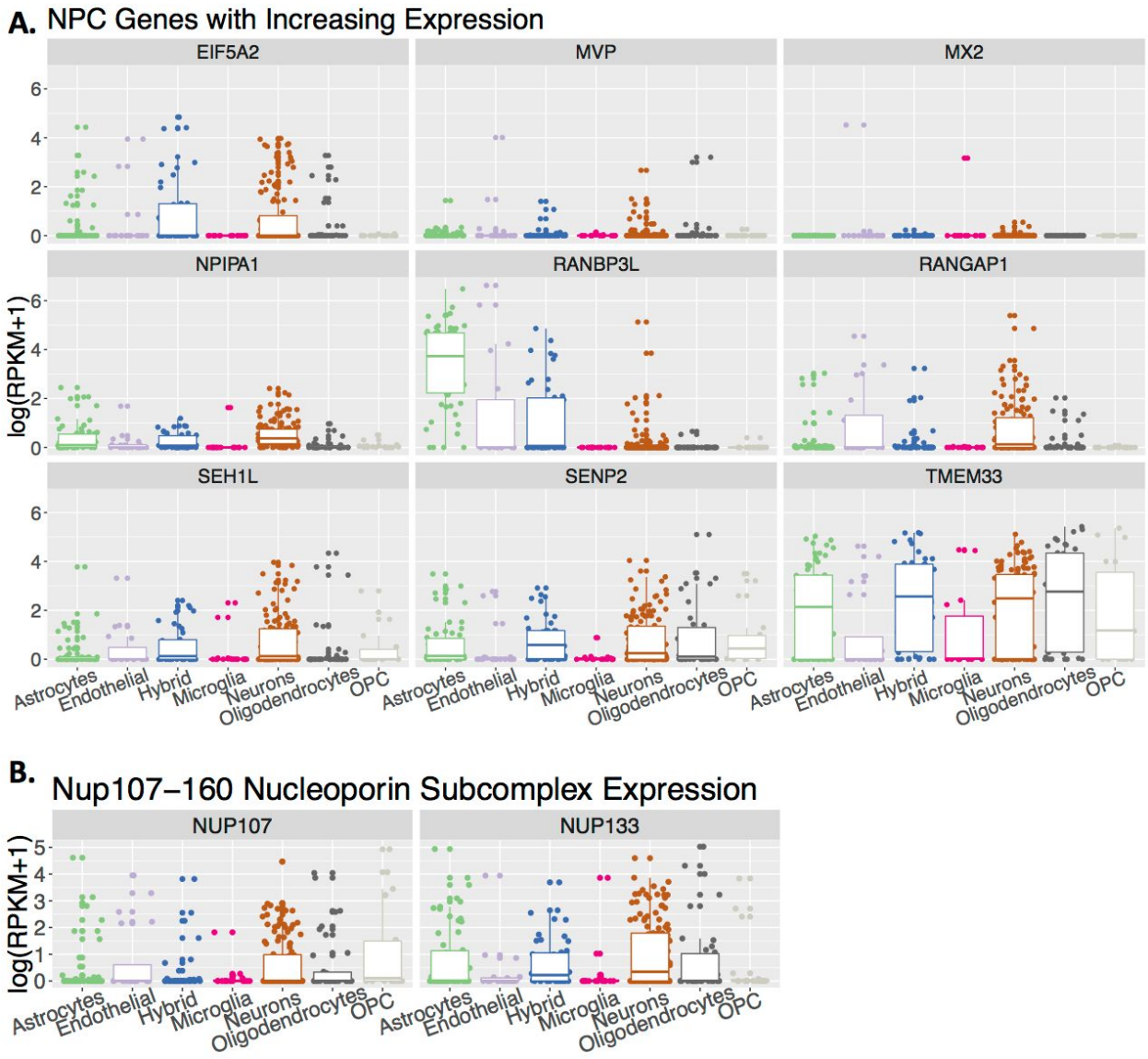

Figure S8: Cell type-specific expression of nuclear pore complex genes with increasing expression as the brain matures. An integral facet of nuclear transport are the nuclear pore complexes (NPCs) that facilitate passage through the membrane. Although NPC structures are extremely long-lasting, often persisting for the lifespan of a cell (Savas et al. 2012), many components of the NPCs can be exchanged according to the requirements of the cell, including during neuronal development (D'Angelo et al. 2012). Several of the NPC factors with increasing expression are involved with the NUP107-160 nuclear pore subcomplex. For instance, *SEH1L* is

increasingly expressed in cortex and encodes a component of the NUP107-160 nuclear pore subcomplex (Zuccolo et al. 2007). *SENP2* is an increasingly expressed gene that encodes a SUMOylation protease that also interacts with the NUP107-160 nucleoporin subcomplex in addition to other NPC components (Goeres et al. 2011). *RANGAP1* and *RANBP2* are increasingly expressed in cortex, and RANGAP1-RANBP2 is recruited by the NUP107-160 nuclear pore subcomplex to the kinetochores during mitosis (Zuccolo et al. 2007). **(A)** Log[reads per kilobase per million mapped (RPKM) + one] of the nine genes with significantly increasing expression (FDR≤0.05) over cortical development as measured in a single cell RNA-seq dataset of cells isolated from adult human postmortem brain (Darmanis et al. 2015). Although the association of the NUP107-160 nuclear pore subcomplex and RANGAP1 with proliferative phenotypes and mitosis (Vorpahl et al. 2014) likely points to increasing expression due to increasing gliogenesis over the human lifespan, *RANGAP1* is higher expressed in human neurons than in glial subtypes in this single cell RNA-seq dataset, implicating neurons as well (FDR=2.1×10^{-3}). **(B)** log(RPKM+1) as measured in a single cell RNA-seq dataset of cells isolated from adult human postmortem brain (Darmanis et al. 2015). *NUP107* and *NUP133*, genes encoding the main components of the NUP107-160 nuclear pore subcomplex, are not expressed in a cell type-specific way. Although NPC structures are stable in postmitotic neurons of adult cortex and developmental expression of these pore related genes did not explain other results, increasing expression of genes encoding NUP107-160 subcomplex components and interactors may reflect additional roles of the subcomplex in these cells.

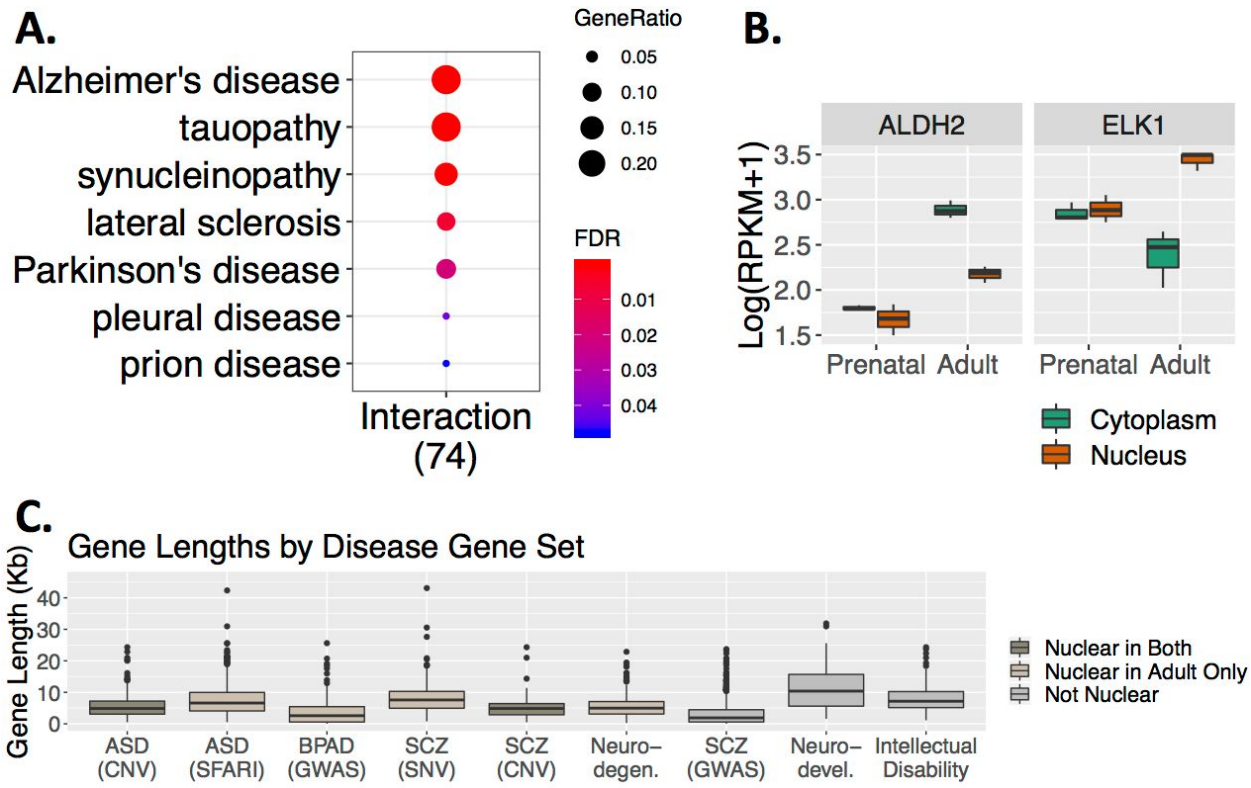

Figure S9: Disease Semantic and Ontology Enrichment. (A) Enrichment for disease ontology terms in "Interaction" genes whose expression varies by both fraction and age. Genes with a significant interaction between subcellular localization and age were enriched for involvement in Alzheimer's disease and other neurodegenerative diseases (abs($\log_2$ fold change)≥1; FDR≤0.05). **(B)** *ALDH2* and *ELK1* gene expression as measured in logarithm of reads per kilobase per million mapped reads plus one read (log(RPKM+1)), grouped by age (adult or prenatal) and fraction (cytoplasm and nucleus). Since the subcellular compartments are globally more similar in prenatal than adult samples, many of the "Interaction" genes were simply more highly expressed in adult than prenatal cortex overall, with greater expression in adult cytoplasm compared to nucleus and relatively similar expression between prenatal fractions. *ALDH2* is an example of greater expression in adult cytoplasm with muted prenatal compartmental differences. Some genes, however, such as the Alzheimer's disease-associated

gene *ELK1*, exhibited other patterns of interaction between fraction and age. Expression of *ELK1*—which encodes a transcription factor that regulates early action gene expression and is implicated in regulating chromatin remodeling, SRE-dependent transcription, and neuronal differentiation—was increased in adult nuclear RNA compared to the cytoplasm. In mice, Elk-1 protein abundance is tightly regulated by subcellular compartment as overexpression in the cytoplasm can lead to cell death (Besnard et al. 2011). **(C)** Length in kilobases (kb) of genes in 9 gene sets associated with autism spectrum disorder (ASD), schizophrenia (SCZ), bipolar affective disorder (BPAD), neurodevelopmental disease (Neuro-devel.), intellectual disability, and neurodegenerative disease (Neuro-degen.). Where applicable, the type of study from which the gene association with disease derives is listed in parentheses. As genes with neuronal functions as a group are longer than average (Gabel et al. 2015) and longer genes are more abundant in the nuclear compartment, gene length may contribute to enrichment of these gene sets. Colors indicate whether the gene set is significantly enriched for genes higher expressed in nuclear RNA in both ages or in adult cortex only (FDR≤0.05). *TTN*, a near 118 kilobase (kb) gene in the SCZ (SNV) set that was 73.3 kb longer than the second longest disease-associated gene, is not depicted.

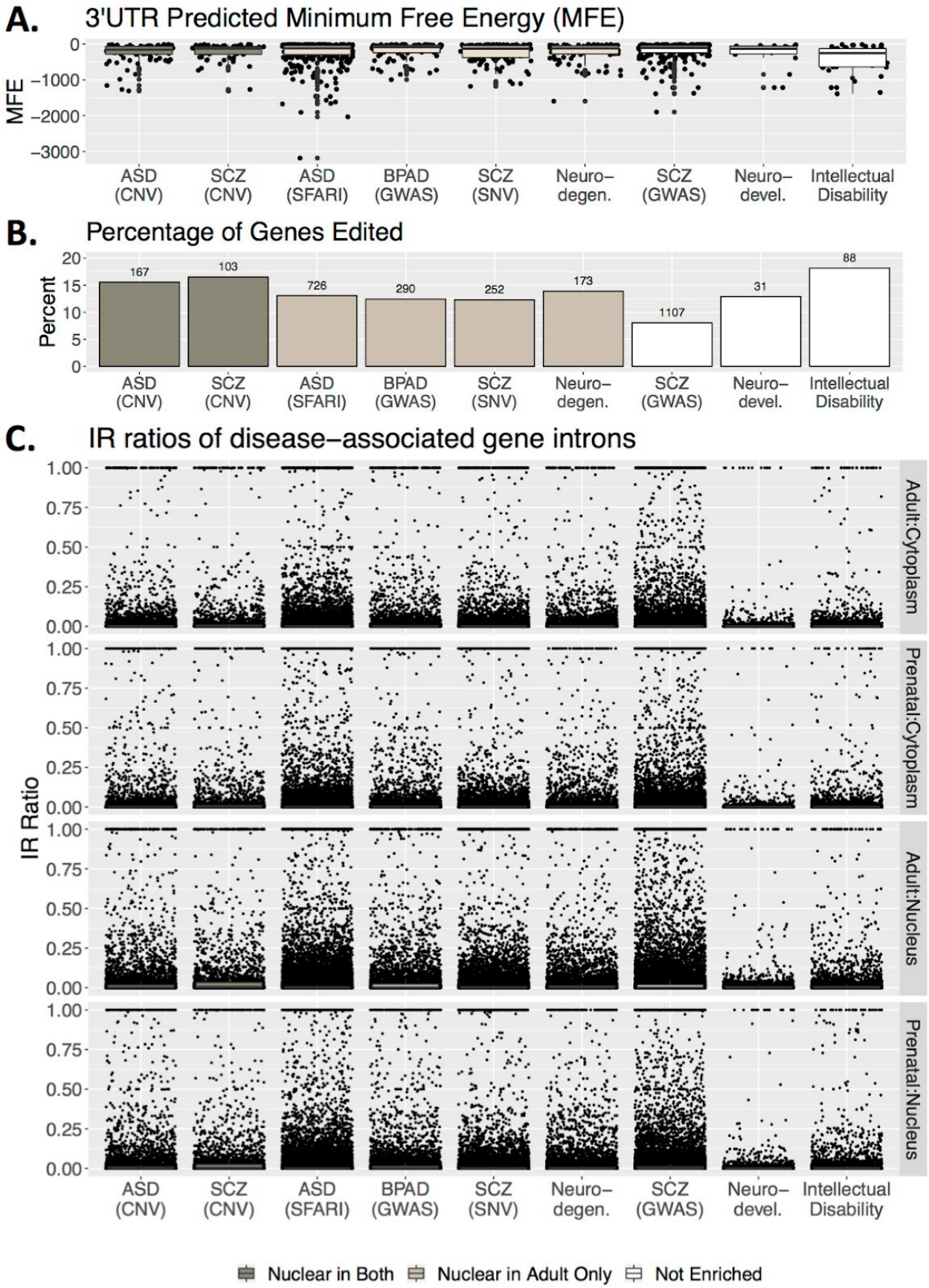
A.
3'UTR Predicted Minimum Free Energy (MFE)
MFE
0
−1000
−2000
−3000
B.
Percentage of Genes Edited
Percent
20
15
10
5
0
167
103
726
290
252
173
1107
31
88
C.
IR ratios of disease−associated gene introns
IR Ratio
1.00
0.75
0.50
0.25
0.00
Adult:Cytoplasm
Prenatal:Cytoplasm
Adult:Nucleus
Prenatal:Nucleus
ASD (CNV)
SCZ (CNV)
ASD (SFARI)
BPAD (GWAS)
SCZ (SNV)
Neuro− degen.
SCZ (GWAS)
Neuro− devel.
Intellectual Disability
Nuclear in Both
Nuclear in Adult Only
Not Enriched

Figure S10: Mechanisms of subcellular localization and disease-associated genes. (A) Minimum predicted free energy of the highest expressed 3'UTR for each gene in each set associated with psychiatric diseases. **B)** The percentage of genes in each set that were found to have an A→I editing site. The total number of genes in each set is listed above each bar. **C)** Intron retention ratio for each intron within a gene associated with psychiatric disease. Gene sets are color-coded according to whether they are enriched for genes that are greater expressed in the nuclear compartment in both adults and prenatal cortex, in only adult cortex, or neither.

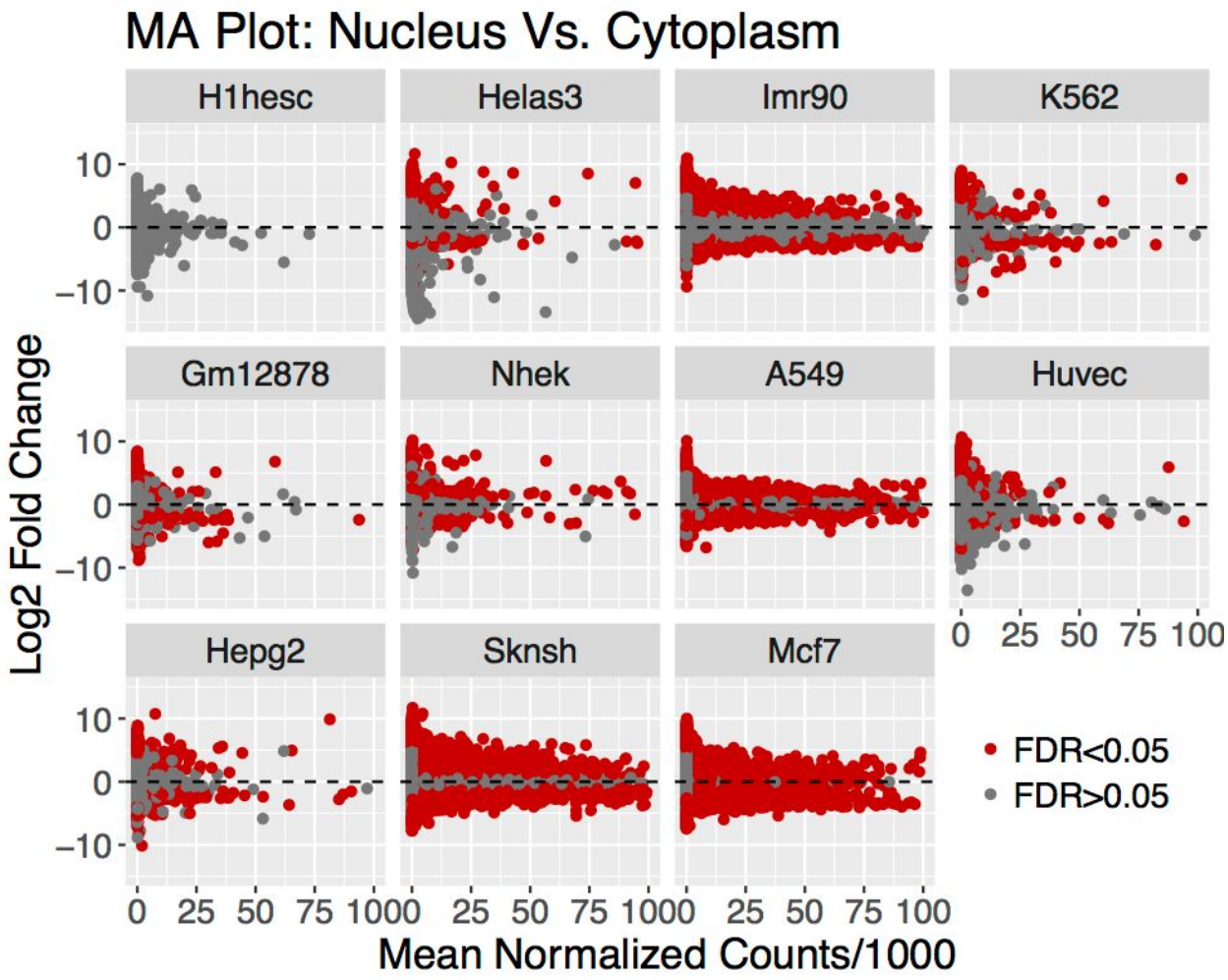

Figure S11: ENCODE Sample Disease Gene Set Association. MA plot showing the $\log_2$ fold change between fractions (positive values indicate greater expression in nucleus) and mean normalized counts divided by 1000 for the 11 cell lines profiled by ENCODE. Red dots indicate FDR≤0.05.

Supplemental References

Besnard A, Galan-Rodriguez B, Vanhoutte P, Caboche J. 2011. Elk-1 a transcription factor with multiple facets in the brain. *Front Neurosci* **5**: 35.

Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR. 2014. Prenatal expression patterns of genes associated with neuropsychiatric disorders. *Am J Psychiatry* **171**: 758–767.

Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F, International Parkinson's Disease Genomics Consortium, 23andMe Research Team, Kerchner GA, Ayalon G, et al. 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* **49**: 1511–1516.

Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, Quake SR. 2015. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci USA* **112**: 7285–7290.

D'Angelo MA, Gomez-Cavazos JS, Mei A, Lackner DH, Hetzer MW. 2012. A change in nuclear pore complex composition regulates cell differentiation. *Dev Cell* **22**: 446–458.

Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, Hemberg M, Ebert DH, Greenberg ME. 2015. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**: 89–93.

Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M, Moran JL, Purcell SM, Sklar P, Sullivan PF, et al. 2016. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat Neurosci* **19**: 1433–1441.

Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrion E, Diallo O, et al. 2011. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* **43**: 860–863.

Goeres J, Chan P-K, Mukhopadhyay D, Zhang H, Raught B, Matunis MJ. 2011. The SUMO-specific isopeptidase SENP2 associates dynamically with nuclear pore complexes through interactions with karyopherins and the Nup107-160 nucleoporin subcomplex. *Mol Biol Cell* **22**: 4868–4882.

Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, Rippey C, Shahin H, Consortium on the Genetics of Schizophrenia (COGS), PAARTNERS Study Group, et al. 2013. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**: 518–529.

International Schizophrenia Consortium. 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**: 237–241.

Itsara A, Wu H, Smith JD, Nickerson DA, Romieu I, London SJ, Eichler EE. 2010. De novo rates and selection of large copy number variation. *Genome Res* **20**: 1469–1481.

Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, International Schizophrenia Consortium, Wellcome Trust Case Control Consortium, Craddock N, Owen

MJ, et al. 2009. Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* **18**: 1497–1503.

Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, Moran J, Chambert K, Toncheva D, Georgieva L, et al. 2012. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* **17**: 142–153.

Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, et al. 2019. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet* **51**: 414–430.

Levinson DF, Duan J, Oh S, Wang K, Sanders AR, Shi J, Zhang N, Mowry BJ, Olincy A, Amin F, et al. 2011. Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. *Am J Psychiatry* **168**: 302–316.

Levy D, Ronemus M, Yamrom B, Lee Y, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, et al. 2011. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**: 886–897.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, et al. 2008. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**: 477–488.

Moeschler JB, Shevell M, American Academy of Pediatrics Committee on Genetics. 2006. Clinical genetic evaluation of the child with mental retardation or developmental delays. *Pediatrics* **117**: 2304–2316.

Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, Dominov JA, Kenna BJ, Nalls MA, Keagle P, et al. 2018. Genome-wide analyses identify *KIF5A* as a novel ALS gene. *Neuron* **97**: 1268-1283.e6.

Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, et al. 2018. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**: 381–389.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, et al. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**: 368–372.

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, et al. 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**: 863–885.

Savas JN, Toyama BH, Xu T, Yates JR, Hetzer MW. 2012. Extremely long-lived nuclear pore

proteins in the rat brain. *Science* **335**: 942.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, et al. 2007. Strong association of de novo copy number mutations with autism. *Science* **316**: 445–449.

Simons Foundation Autism Research Initiative. 2019. SFARI Gene. https://www.sfari.org/resource/sfari-gene/ (Accessed October 6, 2019).

Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, Mattheisen M, Wang Y, Coleman JRI, Gaspar HA, et al. 2019. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**: 793–803.

Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, et al. 2008. Large recurrent microdeletions associated with schizophrenia. *Nature* **455**: 232–236.

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, et al. 2012. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**: 525–537.

Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, Zhang R, Ramaswami G, Ariyoshi K, et al. 2017. Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**: 249–254.

Vacic V, McCarthy S, Malhotra D, Murray F, Chou H-H, Peoples A, Makarov V, Yoon S, Bhandari A, Corominas R, et al. 2011. Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. *Nature* **471**: 499–503.

Vorpahl M, Schönhofer-Merl S, Michaelis C, Flotho A, Melchior F, Wessely R. 2014. The Ran GTPase-activating protein (RanGAP1) is critically involved in smooth muscle cell differentiation, proliferation and migration following vascular injury: implications for neointima formation and restenosis. *PLoS ONE* **9**: e101519.

Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, et al. 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**: 539–543.

Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, Levy S, Gogos JA, Karayiorgou M. 2012. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet* **44**: 1365–1369.

Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* **40**: 880–885.

Zuccolo M, Alves A, Galy V, Bolhy S, Formstecher E, Racine V, Sibarita J-B, Fukagawa T, Shiekhattar R, Yen T, et al. 2007. The human Nup107-160 nuclear pore subcomplex contributes to proper kinetochore functions. *EMBO J* **26**: 1853–1864.