**Title**: Characterizing the nuclear and cytoplasmic transcriptomes in developing and mature human cortex reveals changing RNA regulation patterns

**Authors**: Amanda J. Price, Taeyoung Hwang, Ran Tao, Emily E. Burke, Anandita Rajpurohit, Jooheon Shin, Thomas Hyde, Joel E. Kleinman, Andrew E. Jaffe, Daniel R. Weinberger

**Abstract** (~200 words)

**Introduction**

Human brain development is characterized by precisely-timed changes to gene expression across the lifespan, particularly at the prenatal to postnatal transition1–3. One level of regulation used is the compartmentalization of the whole transcriptome by the nuclear membrane.

Many factors influence the composition of nuclear and cytoplasmic transcriptomes. For instance, as most splicing of pre-mRNA to mature mRNA occurs co-transcriptionally in the nucleus4,5, this leads pre-mRNA and longer genes that take more time to be transcribed and exported to be overrepresented in the nucleus compared to cytoplasm6–8. Recent studies have also highlighted the role of the nuclear membrane as a transcriptional noise buffer, filtering stochastic bursts of gene expression from the cytoplasm by retaining mature mRNA transcripts in the nucleus9,10. Nuclear retention can also regulate the timing of cytoplasmic activity of a transcript11,12 as well as perform quality control by sequestering aberrant transcripts in the nucleus and targeting them for degradation.

The mechanisms by which import and export decisions are made across the nuclear membrane are myriad and an area of active research. Interestingly, many of these RNA trafficking mechanisms are particularly prevalent in brain, and have been shown to play a role in development. For instance, alternative splicing—particularly intron retention—has been shown to regulate RNA localization as a mechanism to suppress lowly expressed transcripts13,14. This suppression can occur via nuclear retention by nuclear surveillance machinery and degradation by exosomes, or via nonsense mediated decay (NMD) in the cytoplasm. Intron retention is highly prevalent in neural lineages and serves to down-regulate genes involved in other lineage fates during neural differentiation15–17. RNA editing has also been shown to be developmentally regulated in human brain, with a subset of editing sites associated with neuronal maturation18. In at least one example RNA editing has also been shown to regulate activity-dependent nuclear transcript retention, although global characterization of RNA editing patterns by subcellular fraction shows RNA editing not to be broadly necessary for nuclear retention11,19. Recent studies have also shown that disruption of proper nucleocytoplasmic transport of proteins and RNA plays a role in aging as well as neurodegenerative disorders such as fronto-temporal dementia and amyotrophic lateral sclerosis20,21. Given this accumulation of evidence, subcellular RNA localization may therefore play an underappreciated role in the etiology of developmental brain disorders. Although nuclear and cytoplasmic transcriptomes have been assessed using *in vitro* models, subcellular fractions have not yet been characterized in cortical brain tissue.

Characterizing the nuclear and cytoplasmic fractions in human brain has additional relevance given the increasingly frequent use of nuclear RNA in single cell and cell population-based studies of human brain22,23. Because frozen post-mortem brain tissue is difficult to dissociate to a single cell suspension, most protocols instead isolate nuclei, which are more robust to degradation24. In transcriptomic studies, this limits analysis to the nuclear RNA fraction, therefore understanding the differences in composition between compartments over human brain development will help inform future studies using nuclear RNA without a comparable cytoplasmic fraction.

To address these questions in human cortical tissue, we characterized the nuclear and cytoplasmic transcriptomes in developing and mature prefrontal cortex using two RNA sequencing library preparation methods and examined distributions of gene sets associated with neurodevelopmental and neurodegenerative disorders. We show that although many genes are differentially expressed by fraction, developmental differences in gene expression are similarly detectable in nuclear and cytoplasmic RNA. Interestingly, differentially localized genes by fraction are enriched for not only neurodegenerative disease but also autism spectrum disorder gene sets. We confirm intron retention to be the splice variant type most differentially utilized by fraction, and identify several introns differentially regulated by fraction and age. Finally, we characterize RNA editing patterns across fraction and age, and identify groups of editing sites that are unique to the nucleus. <last sentence summarizing RBP results and RNA secondary structure, editing sites in splice sites?>

**Results**

As described in Fig. S1, we sequenced nuclear and cytoplasmic RNA isolated from three prenatal and three adult human brains. Because total RNA from a given sample is dominated by rRNA, different strategies can be employed to improve the signal of other RNA species in the sequencing data 25–27. “PolyA” library preparation selects polyadenylated transcripts via a pull-down step, while “Ribozero” library preparation relies on a rRNA depletion step. Together, these library methods capture the transcriptomic diversity in these subcellular compartments in developing human brain because of their respective preferences for mature mRNA and unpolyadenylated transcripts (e.g., ncRNA or pre-mRNA), as seen in Fig. S2A. One adult nuclear “Ribozero” sample failed quality control and was discarded. Demographic and sequencing information can be found in Table S1. Two prenatal cytoplasmic “PolyA” samples had higher read depth and were downsampled to a comparable depth. In total 43,610 ensembl genes were expressed across samples.

The quality of fractionation was confirmed by determining that genes known to localize either to the nucleus (i.e., *MALAT1*) or cytoplasm (i.e., *ACTB*, *FMR1*) were significantly enriched in the appropriate compartment (False Discovery Rate (FDR)<0.01; Fig. S2B), although prenatal samples showed less enrichment than adult (FDR=1.2e-6 and FDR=9.6e-9 for *ACTB* and *MALAT1* in adult, versus FDR=0.17 and FDR=0.44 in prenatal, respectively).

*Developmental gene expression changes in human cortex are similarly detectable in nuclear and cytoplasmic RNA*

We first defined the RNA content differences between subcellular fractions and replicated many characteristics that have previously been described 4–7,9,28,29. Genes that are significantly more abundant in the nucleus were overall longer than genes more abundant in the cytoplasm, perhaps due to the longer temporal requirement for transcription and passive diffusion through the nuclear pore (Fig. S2C). The proportion of reads aligning to introns was greater in the nucleus than the cytoplasm in both PolyA and Ribozero samples (t>4.7, FDR≤5.96e-3), indicating a greater proportion of immature pre-mRNA transcripts (Fig. S2D). Because Ribozero libraries do not require polyadenylation for sequencing, a greater proportion of differentially expressed genes were non-coding in Ribozero samples (Fig. S3A-B). Genes significantly greater expressed in the nuclear compartment in both adult and prenatal cortex were associated with GTPase binding and protein serine/threonine kinase activity (FDR≤0.05).

However, expression patterns were overall similar between fractions at the gene level. Principal component analysis showed that sample age and library type were the largest contributors to transcriptomic variation, explaining 53% and 35% of the variance (Fig. 1A). Assessing developmental changes in both fractions identified similar numbers of differentially expressed genes (Fig. 1B). Developmental expression trajectories were highly correlated between the fractions (*ρ*=0.89, t = 335.8, p < 2.2e-16; Fig. 1C), and 41-63% of significantly regulated genes overlapped in the four groups (Fig. 1D).

*Prenatal and adult human cortex show distinct patterns of RNA localization across the nuclear membrane*

We next examined the relationship between developmental stage and gene expression by fraction and found that prenatal and adult cortex exhibited unique RNA localization patterns across the nuclear membrane. We identified 1,892-1,894 genes differentially expressed by fraction in adult cortex, but only 30-40 genes differentially expressed in prenatal cortex (Fig. 2A, abs(Log2 Fold Change (LFC)) ≥ 1; FDR ≤ 0.05). This localization pattern difference was also seen in Ribozero samples (Fig. S3C-D). Interestingly, most differentially expressed genes in prenatal samples were more abundant in the nuclear compartment (Table S2). Despite fewer genes being differentially expressed by fraction in prenatal cortex, subcellular expression patterns were correlated between prenatal and adult (*ρ*=0.60, t = 125.9, p < 2.2e-16; Fig. 2B). Although a larger proportion of genes enriched in the nuclear fraction were non-coding than those enriched in the cytoplasm (OR=0.25, p=2.2e-16), the majority (83.5%) of all fraction-regulated genes were protein coding (Fig. 2C).

We then examined the relationship between developmental expression trajectories in groups of genes differentially expressed by fraction and found that the association of fraction expression with developmental expression depended on the fraction in which the developmental changes were measured (Fig. 2D). When measured in cytoplasmic RNA, nuclear-enriched genes in adult cortex had enriched expression in prenatal cortex compared to adult (OR=1.32, FDR=0.013), suggesting that perhaps these RNA products were being sequestered in the nucleus in adult cortex. Cytoplasm-enriched genes in adult cortex, on the other hand, were depleted in prenatal compared to adult (OR= 0.44, FDR = 1.4e-18). When measured in nuclear RNA, however, nuclear-enriched genes in adult cortex were significantly depleted in prenatal and enriched in adult cortex (FDR= 2.01e-10, OR= 0.51), and cytoplasm-enriched genes in adult were neither enriched nor depleted by developmental stage (FDR=1.0, OR= 1.02).

Assessing age-associated expression differences in nuclear and cytoplasmic RNA revealed a complementary perspective to the developmental regulation of transcription by subcellular fraction. Genes significantly differentially expressed by age are listed in Table S3. Genes with increasing and decreasing developmental trajectories when measured in both fractions were on average higher expressed in cytoplasm (Fig. 2E). Interestingly, in adult cortex, genes with developmentally decreasing expression (i.e., greater expression in prenatal than adult cortex) when measured in cytoplasmic but not nuclear RNA were more likely to be higher expressed in nucleus (OR=38.0, FDR=1.9e-20), while those in nuclear but not cytoplasmic RNA were less expressed in nucleus (OR=0.071, FDR=1.7e-10). Likewise, developmentally increasing genes (i.e., greater expression in adult than prenatal cortex) when measured in cytoplasmic RNA were less expressed in nucleus in adult cortex (OR=0.038, FDR=4.7e-24), while those measured in nuclear RNA were more likely to be greater expressed in nucleus than cytoplasm (OR=19.3, FDR=5.5e-10). Taken together, these patterns suggest an inverted relationship between developmental gene expression changes and subcellular compartment RNA localization in that down-regulated genes are more abundant in the nuclear compartment.

*Intron retention patterns in prenatal and adult human cortex associate with mRNA distribution*

Because alternative splicing—particularly intron retention—has been implicated as a mechanism of localization of transcripts within the cell13,14 and can play a role in regulating developmental gene expression15,16, we characterized alternative splicing across the PolyA samples. Unlike in the Ribozero samples, in which the proportion of reads spanning splice junctions was lower in the nuclear than cytoplasmic fraction (t=-4.3, FDR=0.016), the proportion of splice junctions was not significantly different in PolyA samples, as pre-mRNAs were depleted by polyA-selection (t=-1.0, FDR=0.344; Fig. 3A). All following splicing analyses were therefore done using the PolyA samples.

Across all samples, skipped exons (SE) and intron retention (IR) represented the greatest percent of unique splice variants identified (22.7% and 28.1%, respectively; Fig. S4A). 42.8% more unique splice variants were identified in nuclear than cytoplasmic RNA, and 72.9% more in prenatal than adult cortex (Fig. S4B). This is in agreement with prenatal cortex showing a higher proportion of splice junctions than adult cortex (Fig. 3A). Testing differential splice variant use by fraction showed that IR was overrepresented in both adult and prenatal samples (OR>2.5, FDR<2.4e-06). As in gene-level expression, far fewer splicing variants were significantly differentially expressed by fraction in prenatal compared to adult (Fig. S4C). Comparing splice variant expression by age in both nuclear and cytoplasmic RNA, on the other hand, showed the distribution of differentially spliced variant types to be more commensurate with the total proportion of each type (Fig. S4C).

Whether a splice variant was more expressed in nuclear than cytoplasmic RNA, or in prenatal than adult RNA, related to its variant type (Fig. S4D). For instance, significantly differentially expressed IR events by fraction were more likely to be higher expressed in the nucleus (OR=50.9, FDR=8.7e-96) while SE, distal alternative 5’ exon start site (A5SS.D) and 3’ exon start site (A3SS.D) events were more likely to be higher expressed in the cytoplasm (OR=0.091, FDR<1.7e-07). Comparing splicing patterns by age showed that in the cytoplasm, SE, IR, and proximal alternative 3’ start site (A3SS.P) events were more abundant in prenatal than adult cortex (OR>2.3, FDR<2.3e-05). In the nuclear compartment however, IR and A3SS.P events were more abundant in adult than prenatal cortex (OR<0.69, FDR<2.2e-02).

To delve more deeply into IR patterns by fraction and age, introns were conservatively filtered to exclude those that overlapped a known exon or were near a known antisense transcript, those with non-uniform intronic coverage, and those with less than four reads spanning a splice junction involving one of the two exon-intron boundaries or less than four reads supporting intron inclusion at the exon-intron boundaries using IRFinder30. After filtering, 166,661-173,125 introns remained per sample, representing 15,345-15,389 unique genes (Fig. S5A). 152,432 introns were shared between all samples. Across samples, 58.68-85.33% introns were constitutively spliced, 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. Overall, introns in nuclear samples had greater IR ratios than cytoplasmic samples (t=69.5, FDR=0; Fig. 3B). Interestingly, although adult samples overall had greater IR ratios than prenatal samples (t=15.9, FDR=3.7e-56), this differed by fraction: IR was greater in adult samples in nuclear RNA (t=25.9, FDR=1.2e-147), but greater in prenatal samples in cytoplasmic RNA (t= -7.121, FDR=3.2e-12).

We next looked into the relationship between RNA subcellular localization and IR by assigning the maximum IR ratio per sample for each gene with at least one intron passing filtering thresholds. Genes with significantly higher expression in nuclear RNA in both adult and prenatal samples (FDR≤0.05) had overall higher IR ratios than genes enriched in the cytoplasm (t>17.7, FDR<1.8E-65; Fig. 3C). The corollary was also true in that genes containing introns with greater than 10% retention were more likely to be significantly enriched in nuclear RNA than cytoplasmic RNA (FDR≤0.05; OR=1.7, FDR= 1.7e-5; Fig. 3D). In terms of developmental expression trajectories, similar to results from global IR analyses, developmental IR patterns in cytoplasmic and nuclear RNA showed opposite relationships: measured in the cytoplasm, genes with developmentally decreasing expression (FDR≤0.05) had overall higher IR ratios (t=7.63, FDR<3.3e-14), while in the nucleus, genes with increasing expression (FDR≤0.05) had overall higher IR ratios (t= -10.5, FDR<3.4e-25; Fig. S5B). Likewise, genes containing introns with greater than 10% retention were more likely to be significantly enriched in adult than prenatal when measured in nuclear RNA (FDR≤0.05; OR=1.4, FDR= 2.1e-7), but depleted in adult when measured in cytoplasmic RNA (FDR≤0.05; OR=1.4, FDR= 2.1e-7; Fig. S5C).

We then measured differential retention of individual introns using the Audic and Claverie test and identified 35 significantly differentially retained introns (dIRs) by fraction in adult and 6 in prenatal cortex; across developmental stages, we identified 10 dIRs in cytoplasmic RNA and 21 in nuclear RNA (FDR≤0.05, Fig. 3E). dIRs 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<2.2e-25). dIRs more retained in the nuclear compartment were less conserved than the pool of total introns tested (t=-3.4, FDR=6.6e-3) and were depleted for containing repetitive elements (OR=0.034, FDR=1.04e-15). Repeats were also depleted in dIRs that were more retained during prenatal development than during adulthood (OR=0.11, FDR=5.0e-06). dIRs by fraction were bimodally distributed across the length of the gene, while dIRs greater retained in adult than prenatal cortex were significantly closer to the 5’ end of the transcript (t=5.06, FDR=1.5e-05, Fig. S5D).

Examining the expression patterns of genes including dIRs provided insight into the relationship between fraction localization and cortical development. Namely, genes including dIRs by fraction were more likely to also include dIRs by age (OR=88.4, FDR=6.5e-04). Developmental dIRs were more likely to be in genes that were significantly differentially expressed by fraction in both prenatal and adult cortex (OR>5.5, FDR<0.032), while dIRs by fraction in in adult cortex were more likely to be in genes that were significantly differentially expressed by age in both fractions (OR>2.8, FDR<0.03). Specifically, dIRs by fraction (all nuclear-increasing) in adult were depleted in genes greater expressed in prenatal than adult (OR<0.30, FDR<0.005), while dIRs by age were enriched in genes greater expressed in nucleus in nuclear RNA (OR<8.98, FDR<9.8e-4) but were depleted in cytoplasmic RNA when measured in adult (OR=0.089, FDR=0.015, Fig. 3F). [Wrapping up sentence]

*RNA Editing by Fraction and Age*

We next profiled RNA editing across subcellular fractions in prenatal and adult cortex. We identified 3,064-5,840 editing sites per sample, finding 25,051 unique sites across the dataset. Of these, 75.5% were A-to-I edited sites, the most common editing pattern (Appears as A:G or T:C in our sequencing data; Fig. 4A). Of the A-to-I edited sites, 1,025 were shared by all four groups (Fig. 4B). Read coverage was fairly even over all samples at edited sites, with a median coverage of 11-12 reads per site across samples (Fig. S6A). In line with previous reports, annotating the A-to-I editing sites showed that 21.7-33.8% fell within intronic sequence and 37.6-50.8% within 3’UTR sequence by group (Fig. S6B). 40.0-42.0% of A-to-I editing sites overlapped an Alu repeat sequence (Fig. S6C). For all following analyses, we focused on the A-to-I sites.

We compared our editing sites to sites identified in other publications and found that 69% of our 18,907 A-to-I editing sites were also detected by the Genotype-Tissue Expression (GTEx) project 31(Fig. S6D). The largest proportion of editing sites (46.3%) was found in brain compared to other tissues in GTEX (Fig. S6E). 13.8% more novel sites were detected in prenatal cortex than adult cortex, while 43.1% more novel sites were detected in nuclear than cytoplasmic RNA.

To assess the relationship between subcellular localization and age in RNA editing, we first assessed editing rate changes across fraction and age in the 1,025 sites shared among the four groups. As shown in Fig. S6F, the distribution of unadjusted p-values suggested that age but not fraction influenced editing rates. After adjusting for false discovery rate, 81 sites were associated with age, while only 9 were associated with fraction and 6 with interaction between age and fraction. Examining the subcellular localization patterns of a subset of 742 editing sites identified in Hwang et al. (2016) that was increasingly edited in human DLPFC as the brain matured showed that of the 576 identified in our dataset, sites detected in adult but not prenatal cortex were more likely to be developmentally increasingly edited (OR=13.5, FDR=2.5e-25). Increasingly edited sites from Hwang *et al* (2016) were significantly more edited in adult than prenatal cortex (t=12.6, FDR=6.4e-34).

Because most editing sites were found in only one group, we next focused on the unique sites found in all samples of a group. The numbers of A-to-I editing sites unique to each group and found in all samples in the group are listed in Fig. 4C and fully listed in Table S4. Genes containing an editing site in this subset were highly edited, containing significantly more sites than other genes (t>3.25, FDR<0.0069).

As with intron retention, presence of these editing sites also was associated with gene expression. For instance, genes that were significantly greater expressed in adult cortex were enriched for editing sites unique to and present in all adult samples (OR=8.9, FDR=6.3e-19), while genes that were significantly greater expressed in prenatal cortex were enriched for editing sites unique to and present in all prenatal samples (OR=25.9, FDR=2.1e-26, Fig. 4D). These prenatal editing sites were also enriched for ribosomes in KEGG pathway analysis (FDR=0.044). In both adult and prenatal cortex, editing sites that were found in all nuclear but no cytoplasmic samples were more likely to occur in genes that were significantly higher expressed in nuclear RNA than other editing sites (OR>2.9, FDR<2.3e-02, Fig. 4E). Relatedly, annotation of these fraction-associated sites showed that editing sites unique to and in all adult cytoplasmic RNA samples were depleted for occurring in introns (OR=0.169, FDR=0.0098) and enriched for occurring in 3'UTR compared to sites in adult nuclear RNA (OR=3.96, FDR=0.0098). In all groups, 63-100% of edited 3’UTRs were from the major isoform (Fig. S6G). Interestingly, although these editing sites were found exclusively in one group, 86.49-100% of edited introns and 96.55-100% of edited exons are expressed in other groups that do not have the edited site.

Because editing sites may disrupt RNA binding protein (RBP) binding, we also examined the enrichment of RBP motifs overlapping editing sites using RBPMap 32. Although no RBPs were enriched or depleted for the groups of unique editing sites in Fig. 4C, 94 of the 114 RBP motifs tested significantly overlapped an A-to-I editing site (FDR<0.05, Fig. S6H). RBP motif enrichment varied by the annotation of the editing site: while editing sites in CDS and 5’UTR were enriched for not overlapping a RBP motif overall (OR>3.4, FDR<5.6e-41), editing sites in introns were enriched for overlapping motifs of SRSF5, G3BP2, and LIN28A (OR>1.2, FDR<3.08e-02), and editing sites in 3’UTRs were enriched for SRSF3, IGF2BP2, IGF2BP3, and FMR1 (OR>1.15, FDR<4.5e-2). SRSF5 and SRSF3 are splicing factors that are part of the spliceosome; LIN28A is a developmentally associated RBP, while IGF2BP2 and IGF2BP3 are associated with regulating the translation of *IGF2* (RefSeq)*.* FMR1, best known for its role in fragile X mental retardation, is also thought to modulate trafficking of mRNA across the nuclear membrane from the nucleus to the cytoplasm (RefSeq).

*Genes whose expression is regulated by fraction are overrepresented in gene sets associated with psychiatric disease*

We finally performed Disease Ontology (DO) Semantic and Enrichment analysis on the sets of genes differentially expressed by 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(LFC)≥1; FDR≤0.05; Fig. S7A). Since the subcellular compartments are globally more similar in prenatal than adult samples, many of these genes were more highly expressed in adult than prenatal cortex, with greater expression in adult cytoplasm compared to nucleus. *ALDH2* is an example of this expression pattern (Fig. S7B). Some genes however, such as the Alzheimer’s disease-associated *ELK1*, exhibited other patterns of interaction between fraction and age (Fig. S7C). Expression of *ELK1—*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 death33.

We then assessed fraction- and age-associated genes with brain disease gene sets for neurodevelopmental, neurodegenerative, and psychiatric disorders curated from many sources, including genome-wide association, copy number variant, and single nucleotide variant studies34. Interestingly, genes that were greater expressed in the nucleus in both ages were enriched for genes associated with Autism Spectrum Disorder (ASD; OR>4.3, FDR<0.013; Fig S7D), as well as schizophrenia (SZ; OR=5.7, FDR=0.038; Fig S7E). Bipolar Affective Disorder (BPAD) was also associated with genes greater expressed in nuclear RNA in adult cortex (OR=2.4, FDR=0.034). A complete list of gene sets and enrichment can be found in Table S5.

**Discussion** (~1000 words)

The higher expression of many mRNA transcripts in the nucleus may be influenced by contamination with immature pre-mRNA sequence. To reduce the 5’ bias of incompletely transcribed RNA, we measured expression differences limited to the last exon of each gene. Prenatal brain tissue still showed less variation in the nuclear and cytoplasmic transcriptomes in both PolyA and Ribozero samples (Supplemental Figure 3B & 3C). These results suggest that as the brain matures, nuclear retention of RNA becomes a more utilized regulatory strategy in cells of the brain.

Alternatively, retained introns have also been implicated in providing ID sequence that targets transcripts to dendrites in neurons 35.

Many of these mechanisms are facilitated by RNA binding proteins (RBPs), such as PTB1, TDP-43, NOVA, DICER, and ADAR2 16,36–38.

That the nuclear transcriptome can broadly stand in for the total transcriptome has implications for cell type-specific studies in human post-mortem brain that rely on fluorescence-activated nuclear sorting.

* Talk about disease, ALS, APOE intron retention 39
* Talk about Ameur 2011 whole cell differences, Jaffe whole cell difference
* Differences between prenatal and adult brain cells:
  + Proliferation (cell cycle), potency (differentiation), cell type, connectivity
* Profiling in vivo tissue rather than in vitro: pitfalls, but also benefits
* Factors affecting transcript localization across cytoplasm and nucleus:
  + Intron retention
  + 3’ UTR length
  + Transcript length
  + RNA secondary structure
  + Transcript isoform (junctions)
  + RNA-editing
  + miRNA degradation
  + NPC dynamics (density, composition)
  + Sequence motifs and predicted RNA-binding protein motifs
  + Temporal dimension: what is being transcribed (1st) vs. exported (2nd)

**Extended Methods**

*Post-Mortem Brain Samples*

Three prenatal and three adult human postmortem brains were selected from the collection of the Lieber Institute for Brain Development for use in this study. Brains in this collection were acquired, dissected, and characterized as described previously 3,40. 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. Brain material was donated and transferred to the Lieber Institute under an approved 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.

*Cytoplasmic and Nuclear RNA Purification and Sequencing*

A diagram of the study design is included in Fig. S1. Homogenate gray matter from the dorsolateral prefrontal cortex (DLPFC) approximating BA46/9 in adults and the corresponding region of PFC in prenatal samples were used for RNA extraction. To purify cytoplasmic from nuclear RNA, we used the Norgen Biotek Corp. Cytoplasmic and Nuclear RNA Purification Kit (Cat # 21000, 37400) following the manufacturer’s protocol including the optional DNase I treatment. RNA-sequencing libraries were prepared from each RNA fraction using PolyA-selection (“PolyA”; Illumina TruSeq Stranded Total RNA Library Prep Kit, Cat #RS-122-2201) and rRNA-depletion (“Ribozero”; Illumina Ribo-Zero Gold Kit (Human/Mouse/Rat), Cat # MRZG126) protocols to enrich for mRNA species. The resulting 24 libraries were then sequenced on one lane of an Illumina HiSeq 2000; the Illumina Real Time Analysis (RTA) module performed image analysis and base calling, and ran the BCL converter (CASAVA v1.8.2), generating FASTQ files containing the sequencing reads.

*Data Processing and Quality Control*

Raw sequencing reads were mapped to the hg19/GRCh37 human reference genome with splice-aware aligner HISAT2 version 2.0.441, with an average 86.8% alignment rate for PolyA samples and average 92.6% alignment for RiboZero samples. Feature-level quantification based on GENCODE (release 25, lift 37) annotation was run on aligned reads using featureCounts (subread version 1.5.0-p3)42. Exon-exon junction counts were extracted from the BAM files using regtools[T. G. L. McDonnell Genome Institute, regtools, (available at https://regtools.readthedocs.io/en/latest/)] v. 0.1.0 and the `bed\_to\_juncs` program from TopHat243 to retain the number of supporting reads. Annotated transcripts were quantified with Salmon44 version 0.7.2. Finally, alignment/processing metrics and the featureCounts results for genes, exons, exon-exon splice junctions, and annotated transcripts were read in and structured into analyzable matrices using R version 3.3.1. As a quality control check, raw FASTQ files were run through FastQC software45, and all samples passed quality statistics including GC content, adapter content, and overall quality. After pre-processing, all samples passed additional QC checks for alignment rate, gene assignment rate, and mitochondrial mapping rates.

*Gene Expression Analysis*

Principal component analysis was done using the plotPCA() function from the DESeq2 bioconductor package46 (Fig. 1A). Read distribution in Fig. S2D was calculated using the read\_distribution.py function in the RSeQC suite47. Annotation features were assigned in a prioritized order, so that reads overlapping coding (CDS) exons were labeled first, then untranslated (UTR) exons, then introns, and finally intergenic regions.

Gene expression differences were measured using the DESeq2 bioconductor package. Samples were segregated by library, fraction and age and compared using several linear models. Gene expression was first modeled by library type in the 11 nuclear samples using “~ Age + Library” (Fig. S2A). To check localization patterns of known nuclear and cytoplasmic genes, we modeled “~ Age + Fraction” separately in the 12 PolyA and 11 RiboZero samples (Fig. S2B). Adult and prenatal samples from each library separately were assessed for differential gene expression by fraction (“~ Fraction”), while nuclear and cytoplasmic samples from each library were separately assessed by age (“~ Age”), culminating in eight sets of results. These results were reported in Fig. 1B-1C, Fig. 2A-2B, Fig. 2D, and Fig. S3C-S3D.

For subsequent gene expression analyses (i.e., those in Fig. 1D-1E, Fig. 2C-2E, Fig. S2C, and Fig. S3A-S3B), a gene was considered significantly differentially expressed if the absolute value of the log2 fold change (LFC) was greater than or equal to one, and if the false discovery rate (FDR) was less than or equal to 5%.

We subset these genes according to whether they were in agreement across ages if measuring changes in expression by fraction, or across fractions if measuring changes by age, resulting in eight groups (e.g., both nuclear, both cytoplasmic, nuclear in prenatal only, nuclear in adult only, cytoplasmic in prenatal only, cytoplasmic in adult only, nuclear in prenatal but cytoplasmic in adult, and cytoplasmic in prenatal but nuclear in adult for comparison of gene expression by fraction). “Interaction” genes were considered those meeting the above criteria using the model “~ Age + Fraction + Age:Fraction” in the 12 PolyA samples.

Gene and disease ontology enrichments were calculated using the compareCluster() function from the clusterProfiler48 bioconductor package. We used brain disease gene sets from Birnbaum et al. (2014) and calculated enrichment of these genes within the nine groups of fraction-associated genes described in the previous paragraph, only without imposing an LFC threshold of the absolute value of one LFC on the fraction-associated genes.

*Splicing Analysis*

The proportion of splice junctions per sample were calculated by dividing the number of reads overlapping a known or predicted splice junction by the total number of reads.

To characterize splice variant type use across the PolyA samples, we used the SGSeq49 bioconductor package. We first extracted features from the bam files using getBamInfo(), then used analyzeFeatures() to predict and quantify splicing events in each bam based on GENCODE (release 25, lift 37). We finally analyzed and summarized that output using analyzeVariants(), setting the minimum denominator to 10. The number of unique splice variants of each type were counted by extracting the types using variantType(). We calculated differential splice variant use by fraction and age using the DEXSeq50 bioconductor package. In building the DEXSeq dataset, we used the variant IDs as the featureID and the event IDs as the groupID in the DEXSeqDataSet() function. Similarly to the gene-level expression analyses, we subset the 12 PolyA samples by fraction and age and compared differential splice variant expression by fraction using the full model “~ sample + exon + fraction:exon” and the reduced model “~ sample + exon.” We compared splice variant expression by age using the full model “~ sample + exon + age:exon” and the reduced model “~ sample + exon.” We then stratified these results by splice variant type and used Fisher exact test to calculate the enrichment of each type in each fraction and age.

To further assess intron retention in the PolyA samples, we filtered introns from the IRFinder-IR-nondir.txt output of IRFinder run on the Human-hg19-release75 reference for each sample. We excluded introns with the “NonUniformIntronCover” warning and those that had anything but “clean” listed in the GeneIntronDetails output column (i.e., excluding "anti-near", "anti-over", "known-exon+anti-near", "known-exon", and "known-exon+anti-near+anti-over"). Introns were further filtered to exclude introns with fewer than four reads spanning the splice junction or a junction using either the 5’ or 3’ exon-intron boundary, or with fewer than four reads supporting intron inclusion at each exon-intron boundary. To assess the relationship between gene expression and IR, we assigned the maximum IR ratio per sample for each gene from this filtered set of introns and compared IR ratios of genes regulated by fraction and age (FDR≤0.05) using Student’s t-test and Fisher exact test.

To quantify differential retention of individual introns, we subset the samples by fraction and age and filtered the IRFinder-IR-nondir.txt output to create four new lists, first filtering to only include the “clean” introns (from the GeneIntronDetails output column), then filtering constitutively spliced introns by group (i.e., adult, prenatal, nuclear, and cytoplasmic). We then used these new files as input to the analysisWithLowReplicates.pl function from IRFinder to calculate differential intron retention between fraction in prenatal and adult, and by age in nucleus and cytoplasm, using the Audic and Claverie test. We calculated the false discovery rate using p.adjust() and setting the n parameter to the total number of clean, non-constitutively splice introns in each comparison. The relationship between intron retention by fraction and age and gene expression was further examined by comparing counts of each using Fisher exact test.

Intron conservation was tested by extracting per base GERP scores for all “clean” introns from the UCSC Table Browser (hg19), calculating the mean score per intron, and comparing the means of groups of introns using Student’s t-test. Repetitive elements in introns were analyzed by downloading the RepeatMasker track from the UCSC Table Browser (hg19) and finding overlaps using findOverlaps() from the GenomicRanges51 bioconductor package.

*RNA Editing Analysis*

RNA editing sites were called in the 12 PolyA samples as described previously18. We annotated the RNA editing sites to genomic features using the GenomicFeatures51 bioconductor package and a transcription database object built on GENCODE (release 25, lift 37). Overlap with repetitive sequences was assessed using the RepeatMasker track downloaded from the UCSC Table Browser (hg19) by finding overlaps using findOverlaps() from the GenomicRanges51 bioconductor package. We compared the editing sites identified in this study with previously identified editing sites using findOverlaps(). We examined the effect of fraction, age, and the interaction of the two on editing rate in the 1,025 sites present in all samples by first filtering the sites to those with a finite and non-NA logit-transformed editing rate in at least 5 samples and with at least one adult, prenatal, nucleus and cytoplasm represented and then using the model “~ Age + Fraction + Age:Fraction.” We compared the pattern of editing in our dataset of the 576 developmentally increasing editing sites from Hwang et al. (2016) present using Fisher exact test.

We defined the sets of fraction- and age-specific editing sites by sites present in all samples of the listed first group that were not found in the second group. For instance, the “Adult Only” sites were present in all six adult samples but no prenatal samples. We assigned each editing site to the nearest gene using distanceToNearest() from the GenomicRanges package and compared the location of the site by fraction or age with the expression enrichment using the Fisher exact test. We identified KEGG pathway enrichment using compareCluster() for the ten groups of unique editing sites, setting the function to "enrichKEGG." Annotation enrichment for these unique sites was also assessed using the Fisher exact test. To identify the major 3’UTR isoform, we identified which 3’UTR had the highest read coverage per gene.

We used RBPMap32 to quantify the enrichment of RNA binding protein (RBP) motifs overlapping the editing sites. We used a region of 10 bases in both direction of an editing site and the RBPMap database of 114 human RBP motifs as input, and calculated the false discovery rate using p.adjust() and setting the n to 2,155,398, or 114 multiplied by the 18,907 A-to-I editing sites input to RBPMap. The motifs were further filtered to those with FDR≤0.05 that overlap the editing site, and compared to genomic feature annotation enrichment using Fisher exact test.

**References**

1. Kang HJ, Kawasawa YI, Cheng F, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011;478(7370):483-489. doi:10.1038/nature10523.

2. Colantuoni C, Lipska BK, Ye T, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011;478(7370):519-523. doi:10.1038/nature10524.

3. Jaffe AE, Shin J, Collado-Torres L, et al. Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. *Nat Neurosci*. 2015;18(1):154-161. doi:10.1038/nn.3898.

4. Tilgner H, Knowles DG, Johnson R, et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res*. 2012;22:1616-1625. doi:10.1101/gr.134445.111.

5. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature*. 2012;489(7414):101-108. doi:10.1038/nature11233.

6. Zaghlool A, Ameur A, Nyberg L, et al. Efficient cellular fractionation improves RNA sequencing analysis of mature and nascent transcripts from human tissues. *BMC Biotechnol*. 2013;13:99. doi:10.1186/1472-6750-13-99.

7. Solnestam BW, Stranneheim H, Hällman J, et al. Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs. *BMC Genomics*. 2012;13:574. doi:10.1186/1471-2164-13-574.

8. Pandya-Jones A, Bhatt DM, Lin C-H, Tong A-J, Smale ST, Black DL. Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA*. 2013;19(6):811-827. doi:10.1261/rna.039081.113.

9. Bahar Halpern K, Caspi I, Lemze D, et al. Nuclear Retention of mRNA in Mammalian Tissues. *Cell Rep*. 2015;13(12):2653-2662. doi:10.1016/j.celrep.2015.11.036.

10. Battich N, Stoeger T, Pelkmans L. Control of Transcript Variability in Single Mammalian Cells. *Cell*. 2015;163(7):1596-1610. doi:10.1016/j.cell.2015.11.018.

11. Prasanth K V., Prasanth SG, Xuan Z, et al. Regulating gene expression through RNA nuclear retention. *Cell*. 2005;123(2):249-263. doi:10.1016/j.cell.2005.08.033.

12. Mauger O, Lemoine F, Scheiffele P. Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity. *Neuron*. 2016;92(6):1266-1278. doi:10.1016/j.neuron.2016.11.032.

13. Boutz PL, Bhutkar A, Sharp PA. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev*. 2015;29(1):63-80. doi:10.1101/gad.247361.114.

14. Braunschweig U, Barbosa-Morais NL, Pan Q, et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res*. 2014:1774-1786. doi:10.1101/gr.177790.114.

15. Wong JJ, Ritchie W, Ebner OA, et al. Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation. *Cell*. 2013;154(3):583-595. doi:10.1016/j.cell.2013.06.052.

16. Yap K, Lim ZQ, Khandelia P, Friedman B, Makeyev E V. Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev*. 2012;26(11):1209-1223. doi:10.1101/gad.188037.112.

17. Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ. Dynamic integration of splicing within gene regulatory pathways. *Cell*. 2013;152(6):1252-1269. doi:10.1016/j.cell.2013.02.034.

18. Hwang T, Park C-K, Leung AKL, et al. Dynamic regulation of RNA editing in human brain development and disease. *Nat Neurosci*. 2016;19(June):1093-1099. doi:10.1038/nn.4337.

19. Chen L. Characterization and comparison of human nuclear and cytosolic editomes. *Proc Natl Acad Sci*. 2013;110(29):E2741-E2747. doi:10.1073/pnas.1218884110.

20. Mertens J, Paquola ACM, Ku M, et al. Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem Cell*. 2015;17(6):705-718. doi:10.1016/j.stem.2015.09.001.

21. Zhang K, Donnelly CJ, Haeusler AR, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature*. 2015. doi:10.1038/nature14973.

22. Lake BB, Ai R, Kaeser GE, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science (80- )*. 2016;352(6293):1586-1590. doi:10.1126/science.aaf1204.

23. Lacar B, Linker SB, Jaeger BN, et al. Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat Commun*. 2017;8:15047. doi:10.1038/ncomms15047.

24. Grindberg R V, Krishnaswami SR, Novotny M, et al. Using single nuclei for RNA-Seq to capture the transcriptome of postmortem neurons. *Nat Protoc*. 2016;in press(3). doi:10.1038/nprot.2016-015.

25. Li S, Tighe SW, Nicolet CM, et al. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat Biotechnol*. 2014;32(9). doi:10.1038/nbt.2972.

26. Sultan M, Amstislavskiy V, Risch T, et al. Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics*. 2014;15(1):675. doi:10.1186/1471-2164-15-675.

27. Cui P, Lin Q, Ding F, et al. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics*. 2010;96(5):259-265. doi:10.1016/j.ygeno.2010.07.010.

28. Reddy AS, O’Brien D, Pisat N, et al. A Comprehensive Analysis of Cell Type–Specific Nuclear RNA From Neurons and Glia of the Brain. *Biol Psychiatry*. 2016:1-13. doi:10.1016/j.biopsych.2016.02.021.

29. Bhatt DM, Pandya-Jones A, Tong AJ, et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell*. 2012;150(2):279-290. doi:10.1016/j.cell.2012.05.043.

30. Middleton R, Gao D, Thomas A, et al. IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol*. 2017;18(1):51. doi:10.1186/s13059-017-1184-4.

31. Tan MH, Li Q, Shanmugam R, et al. Dynamic landscape and regulation of RNA editing in mammals. *Nature*. 2017;550(7675):249-254. doi:10.1038/nature24041.

32. Paz I, Kosti I, Ares M, Cline M, Mandel-Gutfreund Y. RBPmap: A web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res*. 2014;42(W1):361-367. doi:10.1093/nar/gku406.

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

34. Birnbaum R, Jaffe AE, Hyde TM, Kleinman JE, Weinberger DR. Prenatal expression patterns of genes associated with neuropsychiatric disorders. *Am J Psychiatry*. 2014;171(7):758-767. doi:10.1176/appi.ajp.2014.13111452.

35. Buckley PT, Lee MT, Sul JY, et al. Cytoplasmic Intron Sequence-Retaining Transcripts Can Be Dendritically Targeted via ID Element Retrotransposons. *Neuron*. 2011;69(5):877-884. doi:10.1016/j.neuron.2011.02.028.

36. Ameur A, Zaghlool A, Halvardson J, et al. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol*. 2011;18(12):1435-1440. doi:10.1038/nsmb.2143.

37. Neve J, Burger K, Li W, Hoque M, Patel R. Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Res*. 2016:24-35. doi:10.1101/gr.193995.115.

38. Anantharaman A, Tripathi V, Khan A, et al. ADAR2 regulates RNA stability by modifying access of decay-promoting RNA-binding proteins. *Nucleic Acids Res*. 2017;45(7):gkw1304. doi:10.1093/nar/gkw1304.

39. Xu Q, Walker D, Bernardo A, Brodbeck J, Balestra ME, Huang Y. Intron-3 Retention/Splicing Controls Neuronal Expression of Apolipoprotein E in the CNS. *J Neurosci*. 2008;28(6):1452-1459. doi:10.1523/JNEUROSCI.3253-07.2008.

40. Lipska BK, Deep-Soboslay A, Weickert CS, et al. Critical Factors in Gene Expression in Postmortem Human Brain: Focus on Studies in Schizophrenia. *Biol Psychiatry*. 2006;60(6):650-658. doi:10.1016/j.biopsych.2006.06.019.

41. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;(August 2014). doi:10.1038/nmeth.3317.

42. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi:10.1093/bioinformatics/btt656.

43. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36. doi:10.1186/gb-2013-14-4-r36.

44. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-419. doi:10.1038/nmeth.4197.

45. Andrews S. FastQC: A quality control tool for high throughput sequence data. *Babraham Bioinforma*. 2010:http://www.bioinformatics.babraham.ac.uk/projects/. doi:citeulike-article-id:11583827.

46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12). doi:10.1186/s13059-014-0550-8.

47. Benjamini Y, Speed TP. RSeQC: Quality Control of RNA-seq experiments. *Bioinformatics*. 2012;40(10):e72. doi:10.1093/nar/gks001.

48. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287. doi:10.1089/omi.2011.0118.

49. Goldstein LD, Cao Y, Pau G, et al. Prediction and quantification of splice events from RNA-seq data. *PLoS One*. 2016;11(5):1-18. doi:10.1371/journal.pone.0156132.

50. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. 2012:2008-2017. doi:10.1101/gr.133744.111.Freely.

51. Lawrence M, Huber W, Pagès H, et al. Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol*. 2013;9(8). doi:10.1371/journal.pcbi.1003118.