**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, 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 transition (Colantuoni et al., 2011; Jaffe et al., 2015; Kang et al., 2011). 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 nucleus (Djebali et al., 2012; Tilgner et al., 2012), this leads pre-mRNA and longer genes that take more time to be transcribed and exported to be overrepresented in the nucleus compared to cytoplasm (Pandya-Jones et al., 2013; Solnestam et al., 2012; Zaghlool et al., 2013). 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 nucleus (Bahar Halpern et al., 2015; Battich, Stoeger, & Pelkmans, 2015). Nuclear retention can also regulate the timing of cytoplasmic activity of a transcript (Mauger, Lemoine, & Scheiffele, 2016; Prasanth et al., 2005) 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 transcripts (Boutz, Bhutkar, & Sharp, 2015; U. Braunschweig et al., 2014). 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 differentiation (Ulrich Braunschweig, Gueroussov, Plocik, Graveley, & Blencowe, 2013; Wong et al., 2013; Yap, Lim, Khandelia, Friedman, & Makeyev, 2012). RNA editing has also been shown to be developmentally regulated in human brain, with a subset of editing sites associated with neuronal maturation (Hwang et al., 2016). 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 retention (Chen, 2013; Prasanth et al., 2005). 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 sclerosis (Mertens et al., 2015; Zhang et al., 2015). 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 brain (Lacar et al., 2017; Lake et al., 2016). 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 degradation (Grindberg et al., 2016). 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** (~2800 words)

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 (Cui et al., 2010; Li et al., 2014; Sultan et al., 2014). “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” sample had higher read depth and were downsampled to a comparable sequence depth. In total 43,610 ensembl genes were expressed across the 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 overall (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 (Bahar Halpern et al., 2015; Bhatt et al., 2012; Djebali et al., 2012; Reddy et al., 2016; Solnestam et al., 2012; Tilgner et al., 2012; Zaghlool et al., 2013). 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. S1C). The proportion of reads aligning to introns was greater in the nucleus than the cytoplasm in both polyA and Ribozero samples (FDR≤6e-2), indicating a greater proportion of immature pre-mRNA transcripts (Fig. S1D). 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. S3A-B). Interestingly, most differentially expressed genes in prenatal samples were more abundant in the nuclear compartment. A list of differentially expressed genes by fraction is presented in Table S2. Despite fewer genes’ expression being significantly regulated 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). Because Ribozero libraries do not require polyadenylation for sequencing, a greater proportion of differentially expressed genes were non-coding in Ribozero samples (Fig. S3C-D).

We then examined the relationship between developmental expression trajectories in groups of genes differentially expressed by fraction in adult cortex and found that the direction of fraction enrichment association with developmental expression depended on the fraction in which the developmental changes were measured (Fig. 2D). When measured in cytoplasmic RNA, genes that were greater expressed in the nucleus than cytoplasm 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. In contrast, genes with greater cytoplasmic than nuclear expression in adult cortex were depleted in prenatal cortex compared to adult when measured in cytoplasmic RNA (OR= 0.44, FDR = 1.4e-18). When measured in nuclear RNA, on the other hand, genes with higher expression in adult nuclei were significantly depleted in prenatal and enriched in adult cortex (FDR= 2.01e-10, OR= 0.51). Genes with higher expression in cytoplasm in adults 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. 73.3% of genes with significant developmental expression changes were detected in both the cytoplasmic and nuclear transcriptomes (Fig. 2E). When measured in cytoplasmic RNA, genes with decreasing expression over development (i.e., greater expression in prenatal than adult cortex) were more likely to be greater expressed in nucleus than cytoplasm in adult cortex (OR=38.0, FDR=1.9e-20), while genes with decreasing developmental expression when measured in nuclear RNA were more likely to be less expressed in nucleus than cytoplasm in adult cortex (OR=0.071, FDR=1.7e-10, Fig. 2F). Likewise, when measured in cytoplasmic RNA, genes with increasing expression over development (i.e., greater expression in adult than prenatal cortex) were less expressed in nucleus than cytoplasm in adult cortex (OR=0.038, FDR=4.7e-24), while genes with increasing expression over development measured in nuclear RNA were more likely to be greater expressed in nucleus than cytoplasm in adult cortex (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 cell (Boutz et al., 2015; U. Braunschweig et al., 2014) and can play a role in regulating developmental gene expression (Wong et al., 2013; Yap et al., 2012), 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 in cytoplasmic fraction (FDR=3.1e-3), the proportion of splice junctions was comparable in polyA samples, as pre-mRNAs were depleted by polyA-selection (Fig. 3A).

Across all samples, skipped exons (SE) and intron retention (IR) represented the greatest fraction 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 a 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 showed the distribution of differentially spliced variant types to be more commensurate with the total proportion of each type (Fig. S4C).

The direction of splice variant expression by fraction and age was influenced by variant type (Fig. S4D). For instance, significantly differentially expressed IR events by fraction were more likely to occur 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 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, we used the IRFinder program (Middleton et al., 2017). Introns were conservatively filtered to exclude those that overlapped a known exon or were near a known antisense transcript, and those with low or non-uniform coverage or low splicing at the involved splice junctions. 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 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 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 greater 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-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 (Tan et al., 2017)(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 (Paz, Kosti, Ares, Cline, & Mandel-Gutfreund, 2014). 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.* 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 Alzheimers 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 death (Besnard, Galan-Rodriguez, Vanhoutte, & Caboche, 2011).

We then assessed fraction- and age-associated genes with psychiatric disorder gene sets curated from many sources, including GWAS, copy number variants, single nucleotide variants (Birnbaum, Jaffe, Hyde, Kleinman, & Weinberger, 2014). 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 S4.

**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 (Buckley et al., 2011).

Many of these mechanisms are facilitated by RNA binding proteins (RBPs), such as PTB1, TDP-43, NOVA, DICER, and ADAR2 (Ameur et al., 2011; Anantharaman et al., 2017; Neve, Burger, Li, Hoque, & Patel, 2016; Yap et al., 2012).

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 (Xu et al., 2008)
* 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 (Jaffe et al., 2015; Lipska et al., 2006). 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. Demographic information on the six brains is included in Supplementary Table 1.

*Cytoplasmic and Nuclear RNA Purification and Sequencing*

A diagram of the study design is included in Supplementary Figure 1. 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*

All samples including ENCODE samples were aligned and processed using the Lieber RNA-seq pipeline. Etc…

**References**

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., & Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Structural & Molecular Biology*, *18*(12), 1435–1440. http://doi.org/10.1038/nsmb.2143

Anantharaman, A., Tripathi, V., Khan, A., Yoon, J.-H., Singh, D. K., Gholamalamdari, O., … Prasanth, K. V. (2017). ADAR2 regulates RNA stability by modifying access of decay-promoting RNA-binding proteins. *Nucleic Acids Research*, *45*(7), gkw1304. http://doi.org/10.1093/nar/gkw1304

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., … Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. *Cell Reports*, *13*(12), 2653–2662. http://doi.org/10.1016/j.celrep.2015.11.036

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

Besnard, A., Galan-Rodriguez, B., Vanhoutte, P., & Caboche, J. (2011). Elk-1 a transcription factor with multiple facets in the brain. *Frontiers in Neuroscience*, *5*, 35.

Bhatt, D. M., Pandya-Jones, A., Tong, A. J., Barozzi, I., Lissner, M. M., Natoli, G., … Smale, S. T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell*, *150*(2), 279–290. http://doi.org/10.1016/j.cell.2012.05.043

Birnbaum, R., Jaffe, A. E., Hyde, T. M., Kleinman, J. E., & Weinberger, D. R. (2014). Prenatal expression patterns of genes associated with neuropsychiatric disorders. *American Journal of Psychiatry*, *171*(7), 758–767. http://doi.org/10.1176/appi.ajp.2014.13111452

Boutz, P. L., Bhutkar, A., & Sharp, P. A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes and Development*, *29*(1), 63–80. http://doi.org/10.1101/gad.247361.114

Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., … Blencowe, B. J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Research*, 1774–1786. http://doi.org/10.1101/gr.177790.114

Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R., & Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell*, *152*(6), 1252–69. http://doi.org/10.1016/j.cell.2013.02.034

Buckley, P. T., Lee, M. T., Sul, J. Y., Miyashiro, K. Y., Bell, T. J., Fisher, S. A., … Eberwine, J. (2011). Cytoplasmic Intron Sequence-Retaining Transcripts Can Be Dendritically Targeted via ID Element Retrotransposons. *Neuron*, *69*(5), 877–884. http://doi.org/10.1016/j.neuron.2011.02.028

Chen, L. (2013). Characterization and comparison of human nuclear and cytosolic editomes. *Proceedings of the National Academy of Sciences*, *110*(29), E2741–E2747. http://doi.org/10.1073/pnas.1218884110

Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., … Kleinman, J. E. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*, *478*(7370), 519–23. http://doi.org/10.1038/nature10524

Cui, P., Lin, Q., Ding, F., Xin, C., Gong, W., Zhang, L., … Yu, J. (2010). A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics*, *96*(5), 259–265. http://doi.org/10.1016/j.ygeno.2010.07.010

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–8. http://doi.org/10.1038/nature11233

Grindberg, R. V, Krishnaswami, S. R., Novotny, M., Venepally, P., Lacar, B., Bhutani, K., … Lasken, R. S. (2016). Using single nuclei for RNA-Seq to capture the transcriptome of postmortem neurons. *Nature Protocols*, *in press*(3). http://doi.org/10.1038/nprot.2016-015

Hwang, T., Park, C.-K., Leung, A. K. L., Gao, Y., Hyde, T. M., Kleinman, J. E., … Weinberger, D. R. (2016). Dynamic regulation of RNA editing in human brain development and disease. *Nature Neuroscience*, *19*(June), 1093–1099. http://doi.org/10.1038/nn.4337

Jaffe, A. E., Shin, J., Collado-Torres, L., Leek, J. T., Tao, R., Li, C., … Weinberger, D. R. (2015). Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. *Nature Neuroscience*, *18*(1), 154–161. http://doi.org/10.1038/nn.3898

Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., … Sestan, N. (2011). Spatio-temporal transcriptome of the human brain. *Nature*, *478*(7370), 483–9. http://doi.org/10.1038/nature10523

Lacar, B., Linker, S. B., Jaeger, B. N., Krishnaswami, S. R., Barron, J. J., Kelder, M. J. E., … Gage, F. H. (2017). Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nature Communications*, *8*, 15047. http://doi.org/10.1038/ncomms15047

Lake, B. B., Ai, R., Kaeser, G. E., Salathia, N. S., Yung, Y. C., Liu, R., … Zhang, K. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science*, *352*(6293), 1586–1590. http://doi.org/10.1126/science.aaf1204

Li, S., Tighe, S. W., Nicolet, C. M., Grove, D., Levy, S., Farmerie, W., … Mason, C. E. (2014). Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nature Biotechnology*, *32*(9). http://doi.org/10.1038/nbt.2972

Lipska, B. K., Deep-Soboslay, A., Weickert, C. S., Hyde, T. M., Martin, C. E., Herman, M. M., & Kleinman, J. E. (2006). Critical Factors in Gene Expression in Postmortem Human Brain: Focus on Studies in Schizophrenia. *Biological Psychiatry*, *60*(6), 650–658. http://doi.org/10.1016/j.biopsych.2006.06.019

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

Mertens, J., Paquola, A. C. M., Ku, M., Hatch, E., Böhnke, L., Ladjevardi, S., … Gage, F. H. (2015). Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem Cell*, *17*(6), 705–718. http://doi.org/10.1016/j.stem.2015.09.001

Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J. J.-L., … Ritchie, W. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biology*, *18*(1), 51. http://doi.org/10.1186/s13059-017-1184-4

Neve, J., Burger, K., Li, W., Hoque, M., & Patel, R. (2016). Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Research*, 24–35. http://doi.org/10.1101/gr.193995.115

Pandya-Jones, A., Bhatt, D. M., Lin, C.-H., Tong, A.-J., Smale, S. T., & Black, D. L. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA (New York, N.Y.)*, *19*(6), 811–27. http://doi.org/10.1261/rna.039081.113

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

Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., … Spector, D. L. (2005). Regulating gene expression through RNA nuclear retention. *Cell*, *123*(2), 249–263. http://doi.org/10.1016/j.cell.2005.08.033

Reddy, A. S., O’Brien, D., Pisat, N., Weichselbaum, C. T., Sakers, K., Lisci, M., … Dougherty, J. D. (2016). A Comprehensive Analysis of Cell Type–Specific Nuclear RNA From Neurons and Glia of the Brain. *Biological Psychiatry*, 1–13. http://doi.org/10.1016/j.biopsych.2016.02.021

Solnestam, B. W., Stranneheim, H., Hällman, J., Käller, M., Lundberg, E., Lundeberg, J., & Akan, P. (2012). Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs. *BMC Genomics*, *13*, 574. http://doi.org/10.1186/1471-2164-13-574

Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralser, M., … Yaspo, M.-L. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics*, *15*(1), 675. http://doi.org/10.1186/1471-2164-15-675

Tan, M. H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A. N., … Li, J. B. (2017). Dynamic landscape and regulation of RNA editing in mammals. *Nature*, *550*(7675), 249–254. http://doi.org/10.1038/nature24041

Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. a., Chakrabortty, S., Djebali, S., … Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Research*, *22*, 1616–1625. http://doi.org/10.1101/gr.134445.111

Wong, J. J., Ritchie, W., Ebner, O. A., Selbach, M., Wong, J. W. H., Huang, Y., … Rasko, J. E. J. (2013). Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation. *Cell*, *154*(3), 583–595. http://doi.org/10.1016/j.cell.2013.06.052

Xu, Q., Walker, D., Bernardo, A., Brodbeck, J., Balestra, M. E., & Huang, Y. (2008). Intron-3 Retention/Splicing Controls Neuronal Expression of Apolipoprotein E in the CNS. *Journal of Neuroscience*, *28*(6), 1452–1459. http://doi.org/10.1523/JNEUROSCI.3253-07.2008

Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B., & Makeyev, E. V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes and Development*, *26*(11), 1209–1223. http://doi.org/10.1101/gad.188037.112

Zaghlool, A., Ameur, A., Nyberg, L., Halvardson, J., Grabherr, M., Cavelier, L., & Feuk, L. (2013). Efficient cellular fractionation improves RNA sequencing analysis of mature and nascent transcripts from human tissues. *BMC Biotechnology*, *13*, 99. http://doi.org/10.1186/1472-6750-13-99

Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., … Rothstein, J. D. (2015). The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature*. http://doi.org/10.1038/nature14973