**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** (One subsection per point/figure, ~2800 words)

As described in Figure 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 Figure 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 few genes 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 (Figure 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.

*Alternative splicing patterns across subcellular fraction and age*

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 including skipped exons (SE), skipping of two exons (S2E), intron retention (IR), mutually exclusive exons (MXE), proximal and distal alternative 5’ splice site use (A5SS.P and A5SS.D), proximal and distal alternative 3’ splice site use (A3SS.P and A3SS.D), alternative first exon use (AFE) and alternative last exon use (ALE). 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, SE and IR represented the greatest fraction of unique splice variants identified (22.7% and 28.1%, respectively; Fig. S5A). More unique splice variants were identified in nuclear than cytoplasmic RNA, and a greater number in prenatal than adult samples (Fig. S5B). 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. S5C). Comparing splice variant expression by age in both nuclear and cytoplasmic RNA showed the proportion of differentially spliced variant types to be more commensurate with the total proportion of each type (Fig. S5C).

The direction of splice variant expression by fraction and age was influenced by variant type (Fig. S5D). For instance, significantly differentially expressed IR events were more likely to occur in the nucleus (OR=50.9, FDR=8.7e-96) while SE, A5SS.D and 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 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, 13,763-50,164 introns remained per sample, representing 4,254-7,862 unique genes (Fig. S5E). 3,889 introns were shared between all four groups. Across samples, 37.70- 64.67% introns are constitutively spliced; 34.42-56.11% have 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 = 59.9, p < 2.2e-16; Figure 4A). Interestingly, although adult samples overall had greater intron retention than prenatal samples (t = 4.8165, p = 7.3e-07), this differed by fraction: intron retention was greater in adult samples in nuclear RNA (t = 11.717, p < 2.2e-16), but greater in prenatal samples in cytoplasmic RNA (t = -15.03, p < 2.2e-16).

We next looked into the relationship between gene-level RNA localization and intron retention 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 = 31.185, p < 2.2e-16; Figure 4B). The corollary was also true in that genes containing introns with greater than 50% retention were more likely to be significantly enriched in nuclear RNA than cytoplasmic RNA (FDR≤0.05; p = 2.59e-09, fisher exact test; Figure 4C). In terms of developmental expression trajectories, globally there is no clear relationship between intron retention and being significantly differentially expressed by age in either cytoplasmic or nuclear RNA (Supplementary Figure 6A). However, in cytoplasmic RNA, genes with greater expression in prenatal cortex had overall higher proportion of intron retention (t = 9.5913, p < 2.2e-16) than genes enriched in adult cortex; in nuclear RNA, genes with greater expression in adults have greater proportion of intron retention (t = -5.0201, p = 2.6e-07). In genes with a significant interaction between age and fraction, there was a trend toward greater expression of genes with higher IR (IR ratio > 0.5) in prenatal cortex in cytoplasmic RNA (p = 0.134, fisher exact test), and a less pronounced trend toward greater adult expression of high IR genes in nuclear RNA (p = 0.182, fisher exact test; Figure 4D).

We then measured differential retention of individual introns by fraction and age, further filtering the introns to a list of 1,407 measured by fraction and 1,145 measured by age. Introns that were differentially retained by fraction or age shared certain characteristics: they tended to be single rather than clustered within a gene (p<0.05, fisher exact test), and were significantly shorter than the full list of measured introns (t = -5.1573, p < 3.3e-07). No difference was found in mean per-base sequence conservation between groups of introns. Intron retention influenced overall gene level expression patterns; for instance, in both adult and prenatal cortex, introns retained by fraction were more likely to be in genes whose expression also localized by fraction (FDR<0.05; p < 0.01, fisher exact test). Particularly in prenatal cortex, genes containing introns differentially retained by fraction are more likely to be higher expressed in nuclear RNA (p = 0.003621, fisher exact test), although overall they were not more likely to have a higher IR ratio in nuclear than cytoplasmic RNA than non-significantly retained introns. By age, in the cytoplasm, differential intron retention was more likely to occur in genes whose expression also differed by age (p = 0.03199, fisher exact test); there was a trend for this to be the case in nuclear RNA as well (p = 0.0935, fisher exact test). Adult-enriched introns in both nuclear and cytoplasmic RNA have greater expression in prenatal cortex (t = 5.2669, p = 2.089e-05; t = 6.9574, p = 1.356e-10).

Individual intron retention patterns were also distinct between fraction and age. Comparing intron retention across cellular fractions, nuclear-enriched introns are closer to the 3’ end of a transcript than cytoplasm-enriched introns in both adult and prenatal cortex (t = 4.5057, p < 0.0004). Age-associated introns, on the other hand, showed no enrichment for position in a transcript.

Individual intron analysis by age revealed both shared and contrasting developmental changes in IR patterns between the nuclear and cytoplasmic transcriptomes. In nuclear RNA, significantly differentially retained introns by age are more likely to be present in adult than prenatal cortex (p = 0. 003416, fisher exact test). The same trend was true in the cytoplasm: although prenatal cortex had a greater number of introns passing filtering than adult, there was a trend toward a greater proportion of significantly differentially retained introns being more abundant in adult cortex (p = 0.1214, fisher exact test). Genes with significantly differentially retained introns by age in nuclear RNA were more likely to be evenly expressed by age than genes without age-associated introns, which are enriched for greater expression in in prenatal cortex (p = 0.0002336, fisher exact test). In cytoplasmic RNA, on the other hand, genes with significantly differentially retained introns by age trended toward greater expression in prenatal than adult than genes without age-associated introns (p = 0.08126, fisher exact test). The opposite pattern existed concerning introns with high IR ratios (ratio >= 0.5): significantly differentially retained introns by age in nuclear RNA were more likely to have a high IR ratio in adult cortex than non-significantly regulated introns (p = 0.003416, fisher exact test), while there was a trend that those in cytoplasmic RNA are more likely to have even representation of high IR ratio across prenatal and adult cortex than non-significantly regulated introns, which are more likely to be higher in prenatal (p = 0.1214, fisher exact test).

We next examined the relationship between retention patterns of individual introns across fraction and age to better understand the interaction of these factors. We found greater than background overlap of combined lists of significantly differentially retained introns by fraction and age (p = 2.972e-05, fisher exact test); this relationship was stronger when limiting to introns measured in both settings (p = 4.229e-13, fisher exact test). In all introns measured, introns differentially retained by fraction in prenatal and by age in cytoplasm were less likely than background to overlap (p = 0.01107, fisher exact test). Differentially retained introns by fraction are not likely to be so in both adult and prenatal cortex (p = 0.001206, fisher exact test). When looking at only shared introns, many groups showed greater than expected overlap of significantly differentially retained introns, including between those in cytoplasm and nucleus, prenatal and nucleus, adult and nucleus, and adult and cytoplasm (p < 0.001, fisher exact test). In terms of which direction the overlapping significantly differentially retained introns were, most fraction comparisons were more retained in the nucleus, although 2 of 13 introns regulated by fraction in adult and age in cytoplasm were enriched in the cytoplasm. The 13 were greater in prenatal cortex, while the 2 cytoplasmic introns were greater in adult (p = 0.009524, fisher exact test).

In terms of ontology, cytoplasm-enriched intron-containing genes were associated with vesicles, endocytosis and phagocytosis, and protein tyrosine kinase activity in adult cortex, and kidney and eye development and cell adhesion molecule binding in prenatal samples (FDR < 0.05); genes with nuclear-retained introns in prenatal are associated with regulation of filopodium assembly (FDR < 0.03).

Because of the potential to act as a localizing barcode, we next looked into prevalence of repetitive elements in the retained introns. Most differentially retained introns did not contain a repetitive element, although the most abundant form present were short interspersed nuclear elements (SINEs; Figure 4E-F). In introns regulated by age, increasing introns by age (i.e., adult-enriched) have more repetitive elements in nuclear RNA, while decreasing introns have more in cytoplasm (Figure 4F). Overall there isn’t a relationship between containing a repeat and if an intron was significantly differentially retained by fraction, although in adult samples, a greater proportion than expected by chance did contain a repeat (p = 0.04651, fisher exact test). By age, differentially retained introns are more likely to contain repeats than introns not associated with age (p = 8.236e-07, fisher exact test). This is true in both cytoplasmic (p = 0.01061) and nuclear RNA (p = 2.84e-05), and is particularly driven by adult-retained introns (p-value = 1.199e-06), although there is a trend in prenatal retained introns (p-value = 0.09126). Prenatal introns are less enriched for repeats than adult introns (p = 0.03196).

We next looked into particular types of repetitive elements to see if one was driving the association. The distribution of L1 or L2 wasn’t related to whether an intron was differentially retained by fraction or age. There didn’t appear to be a relationship between ALUs and whether an intron was retained or not by fraction, but by age, ALUs were underrepresented in introns significantly regulated by age (p-value = 0.004456). This was driven largely by effects in cytoplasmic rather than nuclear RNA (p-value = 0.04077); also, mostly prenatal introns rather than adult introns (p-value = 0.002533). In repeat-containing introns, the proportion of simple repeats was higher in introns significantly retained by fraction (p = 0.006974). This effect was greater in fraction-regulated introns in adult samples (p-value = 0.01385), and primarily in nuclear-retained introns (p = 0.006696). In repeat-containing introns, the proportion of simple repeats is greater in significantly retained introns by age (p = 5.329e-05); this is true in both cytoplasmic and nuclear RNA (p < 0.05). Primarily this is in introns that are retained in adults (p = 0.0001488).

*RNA Editing by Fraction and Age*

We also profiled RNA editing across subcellular fractions in prenatal and adult cortex. 3,064-5,840 editing sites were identified 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; Figure 5A). Of the A-to-I edited sites, 1,025 were shared by all four groups (Figure 5B). Read coverage was fairly even over all samples at edited sites, with median coverage of 11-12 reads per site across samples (Supplementary Figure 7A). In line with previous reports, annotating the A-to-I editing sites showed that 26.5-40.7% fell within intronic sequence and 39.7-53.2% within 3’UTR sequence by group (Figure 5C). For all following analyses, we focused on the A-to-I sites.

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 Supplementary Figure 7B, the distribution of unadjusted p-values suggests that age but not fraction influenced editing rates. After adjusting for false discovery rate, 27 sites were associated with age, while only 3 were associated with fraction and 2 with interaction between age and fraction.

However, most editing sites were found in only one group. Figure 5D shows the number of A-I editing sites unique to each group and the percentage of all A-I editing sites in each group that it represents. Where the editing site was located affected the group in which the site was found; for instance, editing sites in introns were more likely to be found in nuclear RNA in both adult and prenatal samples, while 3’UTR editing sites were relatively enriched for the cytoplasm (p < 2.2e-16, fisher exact test). CDS and 5’UTR editing sites were not associated with RNA fraction. Sites falling outside annotated genes (i.e., “Other”) were more likely to be cytoplasmic, particularly in adult cortex (p = 0.003324, fisher exact test). By age, in both cytoplasmic and nuclear RNA, a higher proportion of intronic and 5’UTR editing sites were unique to prenatal than adult cortex (p = 1.421e-11; p = 0.01212), while 3’UTR editing sites were more abundant in adult than prenatal (p = 7.539e-15). CDS and intergenic editing sites were not enriched for age.

*Genes regulated by fraction are enriched in psychiatric disease gene sets*

Gene ontology of genes significantly regulated by fraction emphasizes their potential compartment-specific roles; for instance, cytoplasm-enriched genes are associated with GO terms involving the cytoplasm and mitochondria (Supplemental Figure 4A). These GO term associations were also found when using only the last exon coverage. Overall, fewer GO terms were enriched in the nuclear genes, perhaps reflecting greater heterogeneity in this gene set (Supplemental Figure 4B).

We next performed Disease Ontology (DO) Semantic and Enrichment analysis on the sets of genes differentially expressed by fraction and age. Interestingly, genes that were greater expressed in the nucleus in both ages were enriched for genes associated with Autism Spectrum Disorder (ASD; A complete table of gene set results in terms of compartmentalization is available in Table 1.

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; Supp. Figure 3A). As the subcellular compartments are globally more similar in prenatal than adult samples, many of these genes are more highly expressed in adult than prenatal, with greater expression in adult cytoplasm compared to nucleus. *ALDH2* is an example of this expression pattern (Supp. Figure 3B). Some genes however, such as the Alzheimer’s disease-associated *ELK1*, exhibit other patterns of interaction between fraction and age (Supp. Figure 3C). 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—is increased in adult nucleus. 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).

**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

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

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

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