**Title**: Characterizing the nuclear and cytoplasmic transcriptomes in developing and mature human cortex uncovers a potential new avenue for psychiatric disease gene dysregulation

**Authors**: Amanda J. Price1,2, Taeyoung Hwang1, Ran Tao1, Emily E. Burke1, Anandita Rajpurohit1, Jooheon Shin1, Thomas Hyde1,3,4, Joel E. Kleinman1,3, Andrew E. Jaffe1,2,3,5,6, Daniel R. Weinberger1,2,3,4,7

Affiliations:

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

**Abstract**

The diverse functions of transcriptome compartmentalization by the nuclear membrane are enacted via mechanisms such as intron retention and RNA editing that frequently occur in brain and development. Comparing fractions in brain can also inform the limits of what can be learned from single nucleus RNA-sequencing projects derived from postmortem brain. Here we have characterized the nuclear and cytoplasmic transcriptomes from homogenate prenatal and adult human postmortem cortex using two sequencing library preparations. We find that although many genes are differentially expressed by fraction, developmental expression changes are similarly detectable in nuclear and cytoplasmic RNA. Across ages, RNA fractions were more similar in prenatal than adult cortex, and higher nuclear expression was associated with down-regulation in the contrasting age. Measuring intron retention showed that differentially retained introns by fraction were more likely to be differentially retained by age. We find that RNA editing was not globally associated with RNA fraction expression, although genes containing sites unique to a compartment in one age were higher expressed in those samples. Finally, we found nuclear-enriched genes were also preferentially enriched in gene sets associated with neurodevelopmental psychiatric diseases. Together these results clarify the role of nuclear compartmentalization in profiling developmental expression in cortex.

**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 mechanism by which these changes are regulated is via the compartmentalization of the transcriptome by the nuclear membrane into nuclear and cytoplasmic fractions. Taking a snapshot of the composition of each RNA compartment captures factors of both chance and purpose at work: for instance, since most splicing of pre-mRNA occurs co-transcriptionally in the nucleus4,5, pre-mRNA and longer genes that take more time to be transcribed and exported are often 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 selected mature mRNA transcripts in the nucleus9,10. Still, 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 RNA 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 brain development. For example, alternative splicing—particularly intron retention—has been shown to regulate RNA localization as a means to suppress lowly expressed transcripts via sequestration by nuclear surveillance machinery followed by exosome degradation, or via nonsense mediated decay (NMD) in the cytoplasm13,14. Intron retention is highly prevalent in neuronal lineages and serves to down-regulate genes involved in other lineage fates during neuronal 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 that RNA editing is not 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 also 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 and not across age dimensions.

Characterizing the nuclear and cytoplasmic RNA 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 current protocols isolate nuclei, which are more robust to degradation24. Because these protocols limit RNA studies to the nuclear fraction, understanding the compositional differences between compartments over human brain development would 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 early developing and mature prefrontal cortex using two RNA sequencing library preparation methods and examined distributions of gene sets associated with neurodevelopmental, neurodegenerative, and psychiatric disorders. We show that although many genes are differentially expressed by cellular fraction, developmental differences in gene expression are similarly detectable in nuclear and cytoplasmic RNA. Interestingly, gene expression was much more similar across fractions in prenatal than adult cortex, and expression patterns suggested a down-regulation potentially via nuclear retention of some prenatally expressed genes in adult. We also explored potential mechanisms of gene expression regulation by fraction and found that intron retention was an abundant splice variant type and that introns that were differentially retained by fraction were more likely to also be differentially retained by age. RNA editing was confirmed not to be globally associated with RNA localization, although editing sites unique to a compartment in a specific age were associated with higher expression in that compartment. Finally, we found nuclear-enriched genes in both prenatal and adult cortex to be enriched in psychiatric disorder gene sets.

**Results**

We sequenced nuclear and cytoplasmic RNA isolated from three prenatal and three adult human brains (Fig. S1, Table S1). 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 data25–27. PolyA+ library preparation selects polyadenylated transcripts via a pull-down step, while Ribo-Zero 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 non-polyadenylated transcripts (e.g., ncRNA or pre-mRNA, Fig. S2A). One adult nuclear Ribo-Zero sample failed quality control and was discarded and two prenatal cytoplasmic PolyA+ samples that had higher read depth were down-sampled to a comparable depth. In total we profiled 43,610 Ensembl genes that were expressed across these eight groups of samples (adult/cytoplasm/PolyA+, adult/nucleus/PolyA+, prenatal/cytoplasm/PolyA+, prenatal/nucleus/PolyA+, and adult/cytoplasm/Ribo-Zero, prenatal/cytoplasm/Ribo-Zero, adult/nucleus/Ribo-Zero, and prenatal/nucleus/Ribo-Zero). The quality of fractionation was confirmed by determining that genes known to localize either to the nucleus (e.g., *MALAT1*) or cytoplasm (e.g., *ACTB*)9 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 described4–7,9,28,29. 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 Ribo-Zero samples (t>4.7, FDR<5.96e-3), indicating a greater proportion of immature pre-mRNA transcripts (Fig. S2D). Genes significantly greater expressed in the nuclear compartment in both adult and prenatal cortex were associated with more GTPase binding and protein serine/threonine kinase activity (FDR≤0.05). While the majority (83.5%) of all fraction-regulated genes were protein-coding (Fig. S2E), 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). Because Ribo-Zero libraries do not require polyadenylation for sequencing, a greater proportion of differentially expressed genes across fractions were non-coding in Ribo-Zero samples (Fig. S3A-B).

It is worth noting that 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, respectively (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 fraction/library 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 substantially different RNA localization patterns across the nuclear membrane. We identified 1,894 genes differentially expressed by fraction in adult cortex, but only 40 genes differentially expressed in prenatal cortex in the PolyA+ samples (Fig. 2A, abs(Log2 Fold Change (LFC))≥1; FDR≤0.05). This localization pattern difference was also seen in Ribo-Zero 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).

The association of fraction with developmental expression changes depended on the fraction in which the developmental changes were measured (Fig. 2C). Nuclear-enriched genes in adult cortex had enriched expression in prenatal cytoplasm compared to adult (OR=1.32, FDR=0.013), while cytoplasm-enriched genes in adult cortex were depleted in prenatal compared to adult cytoplasm (OR=0.44, FDR = 1.4e-18). On the other hand, nuclear-enriched genes in adult cortex were significantly enriched in adult compared to prenatal nuclear RNA (OR=1.96, FDR=2.01e-10), and cytoplasm-enriched genes in adult were neither enriched nor depleted by developmental stage (FDR=1.0, OR= 1.02). These results suggest that prenatally enriched RNA may be sequestered in the nucleus in adult cortex to regulate the translation of these genes.

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. 2D). Interestingly, genes with greater expression in prenatal than adult cytoplasmic but not nuclear RNA (i.e., developmentally decreasing expression in cytoplasm only) were more likely to be higher expressed in adult nucleus (OR=38.0, FDR=1.9e-20), while developmentally decreasing genes in nuclear but not cytoplasmic RNA were less expressed in adult nucleus (OR=0.071, FDR=1.7e-10). In other words, developmentally down-regulated genes in cytoplasm were more abundant in the nuclear compartment in adults, although nuclear expression was still lower in adult than prenatal. Likewise, genes with greater expression in adult than prenatal cytoplasmic RNA were less expressed in adult nucleus (OR=0.038, FDR=4.7e-24), while developmentally increasing genes in nuclear RNA were more likely to be greater expressed in adult 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 where decreasingly expressed genes may be retained in the nucleus.

*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. The proportion of splice junctions was not significantly different between the nuclear and cytoplasmic fractions in the PolyA+ samples, as pre-mRNAs were depleted by polyA-selection (t=-1.0, FDR=0.34; Fig. 3A). This was in contrast to the results in the Ribo-Zero samples, in which the proportion of reads spanning splice junctions was lower (t=-4.3, FDR=0.016). 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). Interestingly, 42.8% more unique splice variants were identified in nuclear than cytoplasmic RNA (Fig. S4B). Whether a splice variant was more expressed in nuclear than cytoplasmic RNA related to its variant type (Fig. S4D): significantly differentially expressed IR events by fraction (FDR≤0.05) 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).

To delve more deeply into IR patterns by fraction, introns were 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. 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 12 PolyA+ samples. Across samples, 58.68-85.33% of filtered introns were constitutively spliced in each sample, and 12.20-34.63% had an IR ratio (i.e., intronic reads divided by total intron and flanking exon reads) of greater than zero but less than five percent. Overall, introns in nuclear samples had greater IR ratios than cytoplasmic samples (t=69.5, FDR=0; Fig. 3B).

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 parameters. 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). Similarly, genes containing introns with greater than 10% retention were more likely to be significantly enriched in nuclear than cytoplasmic RNA (FDR≤0.05; OR=1.7, FDR=1.7e-5; Fig. 3D).

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 (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 less likely to contain repetitive elements (OR=0.034, FDR=1.04e-15). Locations of dIRs by fraction across the length of the gene were bimodally distributed (Fig. S5B).

In terms of developmental splicing patterns, 72.9% more unique splice variants were identified in prenatal than adult samples (Fig. S4B). This is in agreement with prenatal samples showing a higher proportion of splice junctions than adult samples (Fig. 3A). As in gene-level expression, far fewer splicing variants were significantly differentially expressed by fraction in prenatal compared to adult (Fig. S4D). Comparing splicing patterns by age showed that SE, IR, and proximal alternative 3’ start site (A3SS.P) events were more abundant in prenatal than adult cytoplasm (OR>2.3, FDR<2.3e-05), while IR and A3SS.P events were more abundant in adult than prenatal nucleus (OR<0.69, FDR<2.2e-02). Similarly, 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 in nuclear RNA (t=25.9, FDR=1.2e-147), but greater in prenatal in cytoplasmic RNA (t=-7.121, FDR=3.2e-12). Developmental IR patterns in cytoplasmic and nuclear RNA also showed opposite patterns in terms of expression trajectories: 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. S5C).

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 also to include dIRs by age (OR=88.4, FDR=6.5e-04). Across developmental stages, we identified 10 dIRs in cytoplasmic RNA and 21 in nuclear RNA (FDR≤0.05, Fig. 3E). dIRs more retained in adult than prenatal samples were significantly closer to the 5’ end of the transcript (t=5.06, FDR=1.5e-05, Fig. S5D). Developmental dIRs were also more likely to be in genes that were significantly differentially expressed by fraction and vice versa (OR>2.8, FDR<0.032). Interestingly however, the relationship between dIRs and expression was the opposite as expected: nuclear-increasing dIRs in adult cortex were depleted in prenatal-enriched genes (OR<0.30, FDR<0.005; Fig. 3F). Taken together, profiling IR patterns reveals a set of mutually regulated introns by fraction and age.

*Highly edited genes contain RNA editing sites unique to an age/fraction group that associate with expression levels*

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 18,907 A-to-I edited sites, only 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 in each fraction and age tested (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 Genotype-Tissue Expression (GTEx) project30 and found that 69% of our 18,907 A-to-I editing sites were also detected in this dataset (Fig. S6D). The largest proportion of editing sites (46.3%) was found in brain compared to other tissues in GTEX (Fig. S6E). 43.1% more novel editing sites were detected in nuclear than cytoplasmic RNA.

To assess the relationships between subcellular localization and RNA editing, we first assessed editing rate changes across fraction and age in the 1,025 sites shared among the four groups. The distribution of unadjusted p-values suggested that age but not fraction influenced overall editing rates (Fig. S6F). After adjusting for false discovery rate, 81 sites were associated with age, while only 9 were associated with fraction and 6 with an interaction between age and fraction.

Because most editing sites were found in only one age/fraction group (Fig. 4B), we next focused on the sites found consistently in every sample of a group and never in a contrasting group. These A-to-I editing sites are summarized in Fig. 4C and fully listed in Table S4. Genes containing an editing site in this subset were highly edited, with significantly more edited sites than other genes (t>3.25, FDR<0.0069). Interestingly, although these subsets of unique editing sites were found exclusively in one compared to a contrasting group, 86.49-100% of edited introns and 96.55-100% of edited exons were expressed in the contrasting groups excluding the edited site. In other words, the edited sequence was usually expressed in other fractions and ages despite being edited in only one.

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. 4D), raising the possibility that these editing sites help in signaling nuclear sequestration. Relatedly, annotation of these fraction-associated sites showed that editing sites unique to and in all adult cytoplasmic RNA samples were depleted for intronic (OR=0.169, FDR=0.0098) and enriched for 3'UTR sequence compared to sites in adult nuclear RNA (OR=3.96, FDR=0.0098).

We next examined RNA editing patterns by age and found 13.8% more novel editing sites in prenatal than adult samples. Examining the subcellular localization patterns of a subset of 742 editing sites identified in Hwang *et al.* (2016) that were reported to be 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). As expected, the increasingly edited sites from Hwang *et al.* (2016) were significantly more edited in adult than prenatal cortex (t=12.6, FDR=6.4e-34). Genes that were significantly greater expressed in adult samples 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. 4E). Taken together, these editing patterns suggest an association with editing and expression levels.

*Genes differentially expressed 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 the “Interaction” genes were simply more highly expressed in adult than prenatal cortex overall, with greater expression in adult cytoplasm compared to nucleus and even expression between prenatal fractions. *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 death31.

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 (GWAS), copy number variation (CNV), and single nucleotide variation (SNV) studies32. Neurodegenerative disease genes were enriched for genes that were more highly expressed in cytoplasm than in nucleus in adult cortex (OR=4.3, FDR=1.5e-3), while intellectual disability genes were enriched for cytoplasmic genes in both ages as well as those only enriched in adult cytoplasm (OR>2.7, FDR<0.012). Interestingly, genes that were greater expressed in the nucleus in both ages were enriched for genes associated with both Autism Spectrum Disorder (ASD; OR>4.9, FDR<4.0e-3; Fig S7D), as well as schizophrenia (SCZ; OR=6.5, FDR=0.014; Fig S7E). Bipolar Affective Disorder (BPAD) was also associated with genes greater expressed in nuclear RNA in adult cortex (OR=3.1, FDR=1.5e-3). Intellectual disability, neurodevelopmental disorder, and neurodegenerative disorder gene sets were neither enriched nor depleted for these nuclear-expressed genes. A complete list of gene sets and enrichment can be found in Table S5.

Because genes with neuronal functions as a group are longer than average33 and longer genes are more abundant in the nuclear compartment, we checked if the genes in the disease sets that were over-represented in the nuclear-enriched genes were longer than other genes. While genes in the four nuclear-enriched sets (ASD-associated genes from CNV studies, ASD-associated genes from database, SCZ-associated genes from CNV studies, and BPAD-associated genes from GWAS) were significantly longer than all other genes (t=17.6, FDR=3.04e-55), three of the sets were actually significantly shorter than the genes in the unassociated disease sets (t<-2.7, FDR<9.5e-03; Fig. 5). Only the ASD-associated genes curated from databases were significantly longer than the other disease genes (t=3.0, FDR=4.1e-03). This result was probably skewed by the inclusion of *TITIN*, a near 118 kilobase (kb) gene that was 73.3 kb longer than the second longest disease-associated gene.

To explore this result in an expanded cellular context, we compared fraction profiles of cytoplasmic and nuclear RNA from cell lines sequenced by the ENCODE Consortium5. Strikingly, H1, a human embryonic stem cell line, showed far fewer differences by fraction than the more differentiated cell types tested (9 of 51,502 genes differentially expressed at FDR<0.05; Fig. S8). SK-N-SH, a cell line derived from neuroblastoma, was among the most distinct by fraction (13,985 of 51,502 genes differentially expressed at FDR<0.05).

In the ENCODE data, genes with significantly greater nuclear expression when controlling for cell type (FDR<0.05) were enriched for several disease gene sets, including ASD genes from CNV studies and from databases, BPAD genes, SCZ genes from SNV studies, and neurodevelopmental disease (NDD) genes (OR>1.69, FDR<0.03; (Table S6). SCZ genes from CNV studies and GWAS were enriched in both the nuclear and cytoplasmic compartments (OR>1.7, FDR<0.042).

Fraction-regulated genes in individual cell lines were also associated with disease gene sets (Table S6). Interestingly, nuclear-enriched genes in A549, an adenocarcinomic alveolar basal epithelial cell line, were enriched in ASD, SCZ, and BPAD gene sets (OR>1.96, FDR<1.6e-02). Nuclear genes in the lymphoblast cell line Gm12878 and the hepatocyte cell line Hepg2 were also enriched in ASD gene sets (OR>1.75, FDR<0.04). Nuclear genes in Imr90, a lung myofibroblast cell line, were enriched in the BPAD gene set (OR=2.04, FDR=0.033), and nuclear genes in mammary epithelium line Mcf7 and keratinocyte cell line Nhek were enriched in ASD and SCZ gene sets (OR>1.86, FDR<5.6e-03). Like A549, nuclear genes in SK-N-SH were enriched in ASD, SCZ, and BPAD gene sets (OR>2.1, FDR<4.7e-03).

Intellectual disability- and neurodevelopmental disorder-associated genes were only significantly enriched in nuclear genes in Nhek cells (OR>2.1, FDR<0.029). Cytoplasmic genes in Mcf7 and SK-N-SH cells were enriched for intellectual disability and neurodegenerative disease genes (OR>2.1, FDR<0.016). Overall, these results indicate that subcellular localization of genes associated with psychiatric disease is prevalent not just in brain but also in many other cellular contexts.

**Discussion**

Here we have characterized a snapshot of RNA compartmentalization in developing and mature human postmortem cortex. We find that despite the presence of pre-mRNA, the nuclear RNA compartment can be used as an adequate stand-in for the whole transcriptome when focusing on gene-level expression. Both nuclear and cytoplasmic RNA captured similar numbers of differentially expressed genes between developmental stages, and the magnitude of change detected was highly correlated between fractions. The use of PolyA+ library preparation minimizes the difference between subcellular fractions; the proportion of splice junctions detected was comparable between fractions when measured using PolyA+ libraries, but significantly less so in nuclear RNA when measured with Ribo-Zero.

Interestingly, differences in expression between fractions were much more muted in prenatal compared to adult cortex. We identified over 47 times more genes differentially expressed by fraction in adult than prenatal cortex in PolyA+ samples. Transcription has been shown previously to be more widespread in prenatal brain than at more mature time points, with 4.1% of the prenatal genome transcribed compared to 3.1% of the adult genome3. We also show that prenatal samples had a higher proportion of splice junctions, indicating that a greater volume of prenatal transcription is being processed. Given that the cellular composition of prenatal cortex includes a higher proportion of neural progenitor cells and embryonic stem cells and that these immature cells have a more plastic epigenome34, it is tempting to speculate that as the brain matures, nuclear retention of RNA becomes a more utilized regulatory strategy in cells of the brain. This hypothesis is supported by the H1 embryonic stem cell line showing fewer differentially expressed genes by fraction than the other more differentiated cell lines profiled by ENCODE. It has also been shown that nuclear pore composition changes as cells differentiate and mature35, so it may be that nuclear pores and transport mechanisms are less mature in fetal brain and passage from nucleus to cytoplasm is less restricted.

At the gene level, trends in developmental and subcellular compartment expression patterns suggest nuclear sequestration of developmentally down-regulated RNA. For instance, nuclear-enriched genes in adult were greater expressed in prenatal than adult in cytoplasm, but greater expressed in adult than prenatal in nuclear RNA. Genes with changing age expression in only one fraction also showed an inverted preference for fraction localization in that cytoplasmic decreasing genes were greater in adult nuclear RNA and nuclear deceasing genes were greater in cytoplasm. In other words, genes with greater cytoplasmic expression tended to be higher expressed in the age in which they were measured, while nuclear-enriched genes tended to be higher expressed in the opposite age, suggesting that these nuclear genes, many of which are protein-coding, are being down-regulated by RNA not being exported to the cytoplasm for translation. While this pattern must be tested in single cell types to be confirmed, it suggests an added layer of regulation to be considered in the design of next-generation sequencing studies.

IR has been shown recently to be a common splice variant type that increases during development in several cell types including neurons12,15,16. Here we characterize splicing patterns across fractions in prenatal and adult cortex and confirm that although the majority of introns are constitutively spliced, IR is an abundant splice variant type, particularly in nuclear RNA. Like overall gene expression, specific splice variants passed more readily through the nuclear membrane in prenatal cortex than in adult, leading prenatal-enriched genes to have higher IR ratios than adult in cytoplasm, but lower IR ratios than adult in nucleus. It is unclear what purpose these introns can be playing in the cytoplasm, although previous work has identified potential roles such as targeting transcripts to dendrites in neurons or targeting transcripts for degradation via NMD14,36. Another curious finding was that nuclear-enriched dIRs in adult were found preferentially in genes higher expressed in adult rather than in prenatal overall, in contrast to the gene expression results mentioned above. These introns were distributed primarily in the 5’ end of the gene, in contrast to expectations from previous work that retained introns be toward the 3’ end14. Further work in specific cell types or single cells will be required to resolve these relationships; nevertheless, IR did link developmental and compartmental expression trajectories in the data, given that dIRs by fraction were more likely to be differentially retained by age as well.

By profiling RNA editing across fractions and ages, we found that RNA editing was not globally associated with RNA localization by fraction, although we identified many sites that were unique to a fraction in one age that were found in every sample of that fraction and age group. These unique editing site groups were found in genes that were more highly edited than other genes and that were more highly expressed in those samples than in the opposite age or fraction, although almost all of the exons and introns targeted for editing were present in the other fraction or age in question in an unedited form. The limited read depth in the samples, however, challenges the conclusiveness of the RNA editing analysis. Future work that probes the relationship of these unique sites to localization and expression should study specific cell types at greater coverage.

Finally, we found that nuclear-enriched genes were also preferentially enriched in gene sets associated with neurodevelopmental psychiatric diseases but not with other brain diseases. Previous work has identified the importance of proper nucleocytoplasmic transport in brain diseases, particularly neurodegenerative diseases such as fronto-temporal dementia and amyotrophic lateral sclerosis20,21. Genes associated with these and related diseases were associated with increased adult cytoplasmic expression, in line with their important roles in mature cortex. Surprisingly however, we found that genes associated with neurodevelopmental psychiatric diseases like ASD, SCZ, and BPAD were more likely to have higher expression in the nucleus in both ages tested, but particularly in adult where more nuclear sequestration in general was found. This association was not related to neuronal genes being longer and therefore taking more time to leave the nucleus. Curiously, this preference for nuclear localization extended to other immortalized cell types profiled by ENCODE. This result suggests that these genes may be undergoing extra processing or regulation in the nucleus that may make them more vulnerable to dysregulation.

This study is limited by lack of single cell or cell type-specific insight into these patterns. By using human postmortem brain tissue, we trade improved clinical validity for reduced resolution of nucleocytoplasmic expression patterns. As mentioned previously, prenatal and adult cortices are populated by different cell types in different proportions, each with different proliferation, potency, and connectivity patterns that may influence the import-export decisions across the nuclear membrane. Despite having to average the signal across cells and cell types, that we still see this association between nuclear-expressed genes and psychiatric disease genes suggests that further study of this relationship is warranted.

**Materials and 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 3,37. 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 (“Ribo-Zero”; 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. “Br5339C1\_polyA” and “Br5340C1\_polyA” fastq files were downsampled to 24 million total reads to make the read depth more comparable across samples by joining paired read files, randomly shuffling read order while maintaining read pairs, and limiting the new downsampled FASTQ file to the top 12 million read pairs in the file.

*Data Processing and Quality Control*

Raw sequencing reads were mapped to the hg19/GRCh37 human reference genome with splice-aware aligner HISAT2 version 2.0.438, with an average 86.8% alignment rate for PolyA+ samples and average 92.6% alignment for Ribo-Zero 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)39. 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 TopHat240 to retain the number of supporting reads. Annotated transcripts were quantified with Salmon41 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 software42, 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.

We downloaded ENCODE data in FASTQ format from the Gene Expression Omnibus (accession #GSE26284), and processed them using the same pipeline.

*Gene Expression Analysis*

Principal component analysis was done using the plotPCA() function from the DESeq2 bioconductor package43 (Fig. 1A). Read distribution in Fig. S2D was calculated using the read\_distribution.py function in the RSeQC suite44. 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 Ribo-Zero 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. S2E and Fig. S3C-S3D.

For subsequent gene expression analyses (i.e., those in Fig. 1D, Fig. 2C-2D, 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 clusterProfiler45 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, and filtering disease gene sets for those whose gene symbol were not represented in the genes expressed in the dataset.

ENCODE samples were first analyzed using the linear model “~ Cell type + Fraction + Cell type:Fraction” data in FASTQ format from the Gene Expression Omnibus (accession #GSE26284), and individual cell types were analyzed for fraction expression changes using the model “~ Fraction.”

*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 SGSeq46 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 DEXSeq47 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 IRFinder48 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 GenomicRanges49 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 GenomicFeatures49 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 GenomicRanges49 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.

**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. 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.

31. 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.

32. 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.

33. Gabel HW, Kinde B, Stroud H, et al. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature*. 2015. doi:10.1038/nature14319.

34. Jaffe AE, Gao Y, Deep-soboslay A, et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat Neurosci*. 2016;19(1):40-47. doi:10.1038/nn.4181.

35. D’Angelo MA, Gomez-Cavazos JS, Mei A, Lackner DH, Hetzer MW. A Change in Nuclear Pore Complex Composition Regulates Cell Differentiation. *Dev Cell*. 2012;22(2):446-458. doi:10.1016/j.devcel.2011.11.021.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

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

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.