

## **Genetics of cell-type-specific post-transcriptional gene regulation during human neurogenesis**

Nil Aygün<sup>1,2</sup>, Oleh Krupa<sup>1,2</sup>, Jessica Mory<sup>1,2</sup>, Brandon Le<sup>1,2</sup>, Jordan Valone<sup>1,2</sup>, Dan Liang<sup>1,2</sup>, Michael I. Love<sup>1,3</sup>, Jason L. Stein<sup>1,2,4,\*</sup>

1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

2) UNC Neuroscience Center University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

3) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

4) Lead contact

\*Correspondence: [jason\\_stein@med.unc.edu](mailto:jason_stein@med.unc.edu)

### **Summary**

The function of some genetic variants associated with brain-relevant traits has been explained through colocalization with expression quantitative trait loci (eQTL) conducted in bulk post-mortem adult brain tissue. However, many brain-trait associated loci have unknown cellular or molecular function. These genetic variants may exert context-specific function on different molecular phenotypes including post-transcriptional changes. Here, we identified genetic regulation of RNA-editing and alternative polyadenylation (APA), within a cell-type-specific population of human neural progenitors and neurons. More RNA-editing and isoforms utilizing longer polyadenylation sequences were observed in neurons, likely due to higher expression of genes encoding the proteins mediating these post-transcriptional events. We also detected hundreds of cell-type-specific editing quantitative trait loci (edQTLs) and alternative polyadenylation QTLs (apaQTLs). We found colocalizations of a neuron edQTL in *CCDC88A* with educational attainment and a progenitor apaQTL in *EP300* with schizophrenia, suggesting

genetically mediated post-transcriptional regulation during brain development lead to differences in brain function.

## **Keywords**

RNA-editing, alternative polyadenylation, quantitative trait loci, genome-wide association studies, missing regulation, neurogenesis.

## **Introduction**

Genome-wide association studies (GWAS) have detected many genetic loci associated with risk for neuropsychiatric disorders and inter-individual variation in brain structure and other brain related traits<sup>1-6</sup>. The vast majority of brain-relevant trait GWAS loci have been detected within non-coding genomic regions, and do not change protein coding sequence, implying that they may impact traits through gene regulation<sup>7-9</sup>. Expression quantitative trait loci (eQTL) analysis, which statistically tests the effects of genetic variants on gene expression, have been widely studied in adult brain bulk tissue to interpret the function of brain-relevant trait GWAS loci<sup>10-15</sup>. These studies aggregate steady-state expression across all potential isoforms regardless of post-transcriptional expression modulations. While adult post-mortem bulk tissue eQTLs have identified gene regulatory mechanisms for a subset of brain-trait associated loci, many do not colocalize, leaving their gene-regulatory function unknown. This has recently been termed the “missing regulation” problem in the field of functional genomics<sup>16</sup>. One potential solution could be that brain-relevant trait GWAS loci may function as eQTLs detected only under certain conditions including developmental stage, cell-type, or stimuli<sup>16,17</sup>. Consistent with this, eQTL studies performed in fetal brain bulk data<sup>18,19</sup>, and using cell-type-specific approaches during human neurogenesis and adulthood<sup>20</sup> identified novel genetic loci associated with brain-relevant traits that were not detected in bulk post-mortem adult brain tissue. Although context-specificity has been taken into account in these studies, there are still brain-trait associated loci

that are not explained even by context specific eQTLs. While greater power and additional contexts will likely yield additional gene regulatory effects of brain-trait associated loci, an alternative approach is to study genetic variants influencing brain-relevant traits via alteration of other gene regulatory phenotypes beyond overall expression including post-transcriptional regulation<sup>21–23</sup>.

The most commonly studied post-transcriptional modifications, alternative splicing, has been found to be a major contributor to brain-trait variation using both bulk tissue<sup>18,19,24</sup> and specific cell-types<sup>25,26</sup>. Many other post-transcriptional events exist, including RNA-editing and alternative polyadenylation, and their genetic regulation are in general poorly studied, especially during the process of neurogenesis. RNA-editing is nucleotide changes in RNA sequence relative to those encoded by the DNA sequence. RNA editing can alter protein product encoded by mRNAs<sup>27,28</sup>, location of transcripts<sup>29,30</sup>, splicing<sup>31,32</sup>, transcript stability<sup>33,34</sup>, and microRNA binding sequences<sup>35,36</sup>. In humans, adenosine-to-inosine (A-to-I) changes are the most common type of RNA-editing<sup>37,38</sup>. A-to-I editing events largely overlap with Alu repeats in the human genome and are generated when ADAR enzymes bind to double-stranded RNA hairpins generated by inverted repeat Alus (IRAlus)<sup>22,39,40</sup>. A-to-I RNA-editing occurs in the human brain<sup>41,42</sup>, and has been shown to impact neurotransmission and neurodevelopment<sup>27,28,43–45</sup>. RNA-editing dysregulation has been also associated with neurological disorders including schizophrenia<sup>46,47</sup>, autism spectrum disorder<sup>48,49</sup>, major depression<sup>50,51</sup>, epilepsy<sup>52</sup>, and Alzheimer's disease<sup>53</sup>. Genetic regulation of RNA-editing was observed across adult tissues by performing editing quantitative trait loci (edQTL) analysis, where 1.4% of edit sites (within 228 genes) were significantly regulated by at least one genetic variant<sup>35,40,46</sup>. As a potential mechanism for edQTLs, genetic variants can perturb RNA secondary structure, which can lead to differential RNA-editing in nearby regions<sup>22,35,40</sup>. Though adult brain bulk tissues are most commonly used to study edQTLs<sup>35,46,54</sup>, recent studies have shown that RNA-editing increases

from development to adulthood in the human brain<sup>41,55</sup>, and edQTLs that were not observed in adult brain bulk tissue were detected in bulk fetal brain tissue<sup>55</sup>. Despite the discovery of temporal-specific edQTLs, the cell-type specificity of RNA-editing events and their genetic regulation are as yet unknown during human cortical development and may have been masked in previous bulk tissue studies due to heterogeneity in cell-type composition.

Alternative polyadenylation (APA) is another post-transcriptional modification occurring at the 3'UTR or introns of a given gene in which the poly(A) tail of transcribed mRNA is added in different genomic locations<sup>56,57</sup>. This modification can influence a variety of cellular processes including gene expression, RNA stability, localization, translation rate, and inclusion of microRNA target sequences<sup>56,57</sup>. Dysregulated APA has been found in a variety of brain-relevant diseases including Amyotrophic lateral sclerosis<sup>58</sup>, Parkinson's disease<sup>59</sup> and Huntington's disease<sup>60</sup>. Furthermore, the disruption of several APA regulators have been reported in Oculopharyngeal muscular dystrophy<sup>61</sup> and multiple neuropsychiatric disorders<sup>62</sup>. Previous research described the dynamics of alternative polyadenylation (APA) sites during neuronal differentiation and detected that longer 3'UTR isoforms are abundant in neurons compared to progenitors<sup>63-65</sup>. Alternative polyadenylation quantitative trait loci (apaQTL also known as aQTL) analysis can be applied to assess the genetic alteration of alternative polyadenylation site usage<sup>66-69</sup>. Mechanistically, genetic alteration of polyadenylation sites, polyadenylation signal motifs, or motifs of RNA binding proteins (RBPs) regulating APA can impact polyadenylation site usage<sup>66</sup>. apaQTLs performed in human adult brain bulk tissue have identified many genetic loci associated with alternative polyadenylation site usage<sup>68,69</sup>. However, the temporal and cell-type-specific regulation of apaQTLs in the human brain has also as yet remained unexplored.

In this study using an *in vitro* cell-type-specific system model of human neurogenesis, we systematically evaluated genetic regulation of RNA-editing and APA events within

progenitors and neurons across ~80 different donors. We discovered that RNA-editing was more prevalent in neurons compared to progenitors, likely due to higher expression of the genes encoding ADAR2 and ADAR3 proteins. Alternative polyadenylation also showed differences across cell types, where longer 3' UTR isoforms were observed in neurons as compared to progenitors. We found that common genetic variation was associated with both editing and alternative polyadenylation in a cell-type-specific manner. Hundreds of cell-type-specific edQTLs and apaQTLs were identified that were not previously discovered using fetal bulk brain data. We also observed that these post-transcriptional QTLs, as well as sQTLs showed largely independent regulatory mechanisms as compared to eQTLs. Furthermore, they provided additional interpretation of brain-relevant GWAS loci in addition to eQTLs, suggesting that studying post-transcriptional QTLs is required for a comprehensive understanding of genetic regulation of molecular phenotypes impacting brain development.

## Results

### Cell-type-specific RNA-editing during human neocortical differentiation

We utilized an *in vitro* cell-type-specific system recapitulating human neocortical differentiation in which we previously generated cell-type specific expression, splicing, and chromatin accessibility QTLs<sup>26,70</sup>. This QTL dataset includes progenitors ( $N_{\text{donor}} = 84$ ) and their differentiated, labeled, and sorted progeny, neurons ( $N_{\text{donor}} = 74$ )<sup>26</sup>. We identified RNA-editing events by observing sequence variants in multiple RNA-sequencing reads (supported by at least 10 RNA-sequencing reads in at least 85% of donors per cell-type) using REDIttools software<sup>71</sup> (Figure 1A and see Methods section). These variants have never previously been identified as genetic variations in human populations. We detected 562 and 3,707 RNA-editing sites in progenitors and neurons, respectively (Table S1). We found that these edit sites overlapped with 261 and 825 genes in progenitors and neurons, and 44% and 59% of genes showed multiple RNA-editing events in those cell types, respectively.

To validate that these edit sites were not present in genomic DNA, we investigated nucleotides at the genomic position of edit sites within ATAC-seq reads from our previous studies<sup>70,72</sup>. We observed that when any ATAC-seq reads were present at a predicted edit site across all donors, exclusively the unedited allele was present at 94.5% of edit sites, providing confidence that the predicted edit sites were not rare previously undiscovered genetic variants (Figures S1A-B). To evaluate the similarity of RNA-editing sites detected in each cell-type to previously discovered editing events, we examined features including mismatch content and genomic positions of edits. Since the base-pairing features of inosine and guanosine are similar, Adenosine-to-inosine (A-to-I) RNA editing mediated by ADAR is observed as Adenosine-to-guanosine (A-to-G) mismatches in RNA-seq<sup>73</sup>. We observed 90.4% and 97% of mismatches in progenitors and neurons were A-to-G changes (Figure 1B). Most RNA-editing sites in both cell types were overlapped with Alu repeats, also consistent with ADAR mediated RNA-editing (Figure 1C). The majority of RNA-editing was detected within intronic and 3'UTR gene regions, though some was found in the coding sequence (Figure 1D). We also detected that 88% and 87% of RNA-editing sites in progenitor and neurons were also previously identified in either GTEx Cortex<sup>74</sup> or BrainVar<sup>55</sup> data in which whole-genome-sequencing data paired to RNA-seq were available (Figure 1E), showing consistency of data generated here with previously discovered RNA-editing events. We further evaluated a common local sequence motif for RNA-editing, which is 1bp upstream enrichment and 1bp downstream depletion of guanosine<sup>75</sup> among the RNA-edit sites we discovered. We observed that both RNA-editing sites found in GTEx or BrainVar data (previously annotated) and editing sites which were not found in previous datasets (novel) showed highly similar, and expected, motif enrichment in neurons (Figure 1F). However, only previously annotated edit sites showed the expected motif enrichment in progenitors, likely due to the smaller number of edit sites detected in this category (Figure 1F). Supporting that these novel edit sites were not false positives, 97% and 99% of

novel edit sites in progenitor and neurons did not show edited alleles in any read from the ATAC-seq data (Figure S1A). These results provided evidence that RNA-edit sites explored in our study exhibited characteristics of previously identified RNA-editing events.

Consistent with the detection of more RNA-edit sites in neurons than progenitors, we found that Alu editing index (AEI), which is calculated as the global measurement of A-to-G changes within Alu elements<sup>76</sup>, was significantly higher in neurons than progenitors (Figure 2A). As a potential mechanism leading to global editing differences between cell types, we compared expression of genes encoding ADAR1, ADAR2, and ADAR3 enzymes. We detected that the expression of *ADAR1* was slightly higher in progenitors than neurons; whereas both *ADAR2* and *ADAR3* were strongly upregulated in neurons (Figure 2B). Consistent with this, increased *ADAR1* expression was positively correlated with global editing levels (AEI) only in progenitors, but increased *ADAR2* and *ADAR3* expression was specifically positively correlated with global editing levels in neurons (Figure S2A). Higher *ADAR2* expression and lower *ADAR3* expression were previously found in neurons compared to oligodendrocytes in the adult brain<sup>54,77</sup>, but have not previously been evaluated in progenitors. Also, *ADAR3* has been previously considered as having an inhibitory role in RNA-editing in the adult brain<sup>54,77,78</sup>, but our developmental and cell-type-specific system revealed a positive correlation between AEI and *ADAR3* expression within neurons, suggesting ADAR3 has a unique role during development leading to increased editing in immature neurons. We also detected a smaller AEI index and number of edit sites discovered in fetal bulk data compared to neurons, and 94% and 96% of edit sites discovered in progenitors and neurons were not identified by using fetal bulk brain data (Figures 2A and C). We observed that average read depth was  $17.1\text{M} \pm 5.8$  and  $99.8\text{M} \pm 29.8$  in fetal bulk and cell-type-specific RNA-seq data, respectively, so the novel cell-type specific editing events may be driven by either the lower read depth of RNA-seq samples in the fetal bulk data limiting our power to discover RNA-editing events or heterogeneity of cell-types in bulk data. Moreover, consistent

with the hypothesis that editing increased throughout the development, we also observed higher AEI values in fetal bulk brain samples<sup>18</sup> at older gestation weeks as neuronal production increased, consistent with increased editing observed in neurons (Figure S2B). Overall, here, we provide evidence that increased expression of genes encoding ADAR enzymes are likely responsible for cell-type-specific global RNA-editing and the higher number of RNA-editing events in neurons during human neurogenesis.

Next, we evaluated the downstream functions of RNA-editing. To understand if RNA-editing sites that we discovered were also dysregulated in neuropsychiatric disorders, we evaluated enrichment of cell-type-specific edit sites detected in our sample within previously identified disease-related edit sites which were differentially detected between individuals from case and control groups using adult-bulk post-mortem tissue<sup>46,49,79</sup>. We found that cell-type-specific RNA-edit sites present during neurogenesis that were observed in our model system overlapped with disease-related edit sites found in schizophrenia, glioblastoma, Fragile X, and autism. We observed a greater overlap of disease associated editing events with neurons as compared to progenitors, but these overlaps in both progenitors and neurons occurred more than expected by chance (Fisher's exact test, FDR < 0.05). These findings suggest a developmental and cellular origin of RNA-editing dysregulation in neuropsychiatric disorders (Figure 2D).

We also evaluated whether RNA-editing influences other cellular downstream functions during neurogenesis. We performed high-content imaging of 8-week differentiated neuronal cultures and labeled them with markers of neuronal differentiation (TUJ1 labeling was used for early born neurons) and all nuclei (DAPI) with 31 wells measured per donor on average. We implemented an image analysis pipeline to quantify the percentage of cells labeled with TUJ1 as a measure of that donor's neurogenic potential, observing strong differences across donors (compare Donor 18 with Donor 321 in Figure 2E). We found that an editing site within the 3'UTR



of *CEP104* gene in neurons was positively correlated with the number of cells expressing a neuronal marker TUJ1 (Figure 2E, Table S2). *CEP104* gene encodes a ciliary protein and loss-of-function mutations on CEP104 were found in individuals with a neurodevelopmental condition, Joubert syndrome<sup>80–82</sup>. These results indicate that RNA-editing can influence fate decisions even without having an effect on the amino acid sequence of the protein.

### Genetic regulation of cell-type-specific editing via editing quantitative trait loci (edQTL) analysis

To perform cell-type-specific edQTL analysis, we tested the association of genetic variants with edit rate, which was defined as the read counts supporting the edited allele (mainly the G nucleoside) divided by the total read coverage at the edit site, within +/-100 kb from edit sites. We included only variants and edit sites located in the same gene (excluding intergenic variants) because we hypothesized that alterations in mRNA secondary structure alter editing. We controlled population structure and global editing principal components (PCs) as technical confounders. We controlled for one global PC of gene editing in progenitors and the major known technical confounder (FACS sorting) in neurons (see Methods), which were highly correlated to *ADAR1-3* expression (Figure S2C). We implemented a hierarchical multiple comparisons correction method using eigenMT-FDR at 5% as a significance threshold (see Methods). We identified 101 and 517 edSites, which are the edit sites significantly regulated by at least one genetic variant, within 79 and 279 genes in progenitors and neurons, respectively (Figure 3A, Table S3). Primary edSNPs, which are variants showing the most significant association with edSites, showed stronger associations as they were closer to the edit sites (Figure 3B). We also observed that genes harboring these edit sites in neurons were significantly enriched in biological pathways related to neuronal morphology and metabolism; whereas we did not detect any significant enriched biological pathways in progenitors (Figure 3C). To investigate a potential mechanism whereby genetic variants impact RNA-editing, we assessed the RNA-secondary structures where significant and nonsignificant edQTLs were

located. We found that the majority of significant edQTLs were found within the double stranded RNA-secondary structure, stem, which is substrate for ADAR enzymes in both cell types (Figure 3D). Only significant neuron edQTLs were significantly more enriched within stem structure compared to nonsignificant neuron edQTLs. We did not observe an enrichment in progenitor edQTLs within the stem structure, though this was likely driven by fewer edQTLs discovered in progenitors (Figure 3D).

We observed high cell-type specificity of genetic effects on editing, finding that 86% and 97% of edSites in progenitors and neurons were not detected in other cell types. We also found that 97% and 99% of edSites in progenitor and neurons were not detected in bulk fetal brain tissue, showing additional genetic discovery is enabled using a cell-type specific approach (Figure 3A).

#### Cell-type-specific alternative polyadenylation sites during human neocortical differentiation

Another post-transcriptional modification of interest during differentiation is alternative polyadenylation. We identified and quantified alternative polyadenylation (APA) site usage by applying the QAPA method, which allows identification and quantification of 3'UTR isoforms mapped to annotated polyA sites by using RNA-seq data<sup>65,83</sup> (see Methods). After filtering out lowly expressed 3'UTR isoforms, we detected 19,200 and 18,246 3'UTR isoforms corresponding to alternative polyadenylation sites within 7,711 and 7,801 genes in progenitor and neurons, respectively (Table S4). We observed 2.5 and 2.3 different 3'UTR isoforms per gene on average in progenitors and neurons, respectively. Principal component analysis for APA usage showed that progenitors and neurons were distinctly separated, indicating that cell type has a strong influence on 3' UTR isoform usage (Figure 4A). We also observed that genes which play a role in alternative polyadenylation<sup>56</sup> were differentially expressed across cell-types, including higher expression of *FIP1L1* and *RBBP6* in neurons, suggesting their distinct

regulation and function during differentiation (Figure 4B). We next performed differential isoform usage analysis across cell-types, and identified both cell-type-specific 3'UTR lengthening and shortening. We observed that the longest 3'UTR isoform of a gene was upregulated in neurons for 79% of 3'UTRs genes, consistent with the previous observation that longer 3'UTRs are expressed during differentiation whereas less differentiated and proliferative cell types generally express shorter 3' UTRs<sup>65,83–85</sup>. As an example, we found that longer 3'UTR of *CALM1* gene was upregulated in neurons (Figure 4C). *CALM1* protein is a calcium ion sensor<sup>86</sup>, and the neuron specific longer 3'UTR expression of the *CALM1* gene was previously shown in mice and its deficiency led to impaired neuronal function<sup>87</sup>. In summary, longer 3' UTRs were observed in neurons as compared to progenitors, some of which have previously been shown to be functional, and which may be mediated by the increased expression of FIP1L1 and RBBP6 polyadenylation factors.

#### Genetically altered cell-type-specific alternative polyadenylation sites

We performed alternative polyadenylation QTL (apaQTL) analysis by evaluating the association of each 3'UTR isoform with the genetic variants within +/- 25 kb window of 3'UTR start and end sites (see Methods). We identified 423 and 215 primary apaQTLs within 352 and 184 genes in progenitors and neurons, respectively (Figure 4D, Table S5). Primary apaQTLs showed stronger associations in closer proximity to the APA site (Figure 4E). To investigate a mechanism underlying apaQTLs, we searched for significant apaQTLs that change canonical polyadenylation signal sequences. We found that 13.6% and 16.3% of significant apaQTLs in progenitor and neurons, respectively, were within canonical polyadenylation signal (PAS) sequences. As previously reported, we also found that AAUAAA was the most frequent motif among these overlapped PAS motifs altered by significant apaQTL<sup>69</sup>. As an example, we detected an apaQTL for *RPL22L1* gene in both progenitor and neuron cells, and the T allele

which was the AAUAAA motif matching allele was associated with increase in short 3'UTR and decrease in long 3'UTR (Figure 4F). To evaluate cell-type-specificity of apaQTLs, we utilized  $\pi_1$  statistics. We estimated the proportion of progenitor and neuron primary aSNP-APA pairs that were non-null associations ( $\pi_1$ ) in neuron and progenitor apaQTLs as 36.3% and 47.4%, respectively, showing high cell-type specificity of alternative polyadenylation. We next evaluated the overlap of cell-type-specific apaQTLs with fetal brain bulk data, and observed high cell-type specificity of alternative polyadenylation (Figure 4D), again showing that genetic effects on post-transcriptional modifications have greater discoverability within homogeneous cell types.

### Genomic features distinguishing molecular QTLs

Next, we evaluated the genomic features distinguishing different types of molecular QTLs in order to understand their shared or unique regulatory mechanisms. We evaluated this using molecular QTLs previously identified in the same population of neural progenitor cells (expression, splicing, and chromatin accessibility) together with those identified in this manuscript (editing and polyadenylation)<sup>26</sup>. We observed that both primary edQTLs and apaQTLs were more often near transcription termination sites (TTS) as compared to primary eQTLs which were more often near transcription start sites (TSS) in both cell-types (Figures 5A-B). Unlike primary sQTLs, both primary edQTLs and apaQTLs were found less often near splice sites (Figure 5C). Functional genetic variants are often near the molecular entity they regulate.

Furthermore, we also detected that primary edQTLs were found less often within chromatin accessible regions more accessible within that cell type as compared to eQTLs<sup>70</sup> (Figure S3A, left side). Also, primary apaQTLs were found less often within chromatin accessible regions compared to eQTLs in progenitor cells (Figure S3A, left side). On the other hand, both primary edQTLs and apaQTL were enriched more within RNA Binding Protein (RBP)

binding sites<sup>88</sup> than eQTLs in both cell types (Figure S3A, right side). These findings are consistent with the predicted mechanism of action of eQTLs, alterations in regulatory element activity marked by accessible chromatin, as compared to ed/apaQTLs which are likely dependent on alterations in affinity of RNA binding proteins.

Also, we observed that 2% and 1.8% of edSNPs, variants significantly associated with at least one edit site, were also significantly associated with the expression of the gene harboring the edit sites (edGene) in progenitors and neurons, respectively. 5.8% and 5.6% of apaSNPs significantly associated with at least one APA, were also significantly associated with the expression of the gene harboring the 3'UTR (aGene) in progenitors and neurons, respectively. Given the difference in statistical power between different datasets, we also applied  $\pi_1$  statistics<sup>89</sup> to assess the overlap across edQTLs/apaQTLs/eQTLs and sQTLs. We found that the proportion of progenitor and neuron primary edSNP-edGene pairs that were non-null associations ( $\pi_1$ ) in progenitor/neuron eQTLs were 12% and 2%; the proportion of progenitor and neuron primary apaSNP-aGene pairs that were non-null associations ( $\pi_1$ ) in progenitor and neuron eQTLs were 50% and 34.6%, and the proportion of progenitor and neuron primary sSNP-sGene pairs that were non-null associations ( $\pi_1$ ) in progenitor and neuron eQTLs were 46.5% and 40.6% (Figure 5D). Taken together, these findings suggested that genetic regulation of RNA-editing, alternative polyadenylation, and alternative splicing site usage were mainly independent from eQTLs, consistent with the observations previously reported<sup>22,67</sup>.

To understand whether the small subset of eQTLs shared with ed/apaQTLs were causally related, we performed mediation analysis<sup>90,91</sup> for variant-edit-gene and variant-APA site-gene triplets. While we did not detect any variant-edit-gene triplets supporting the causal forward model; we detected 335 and 49 variant-APA site-gene triplets. As an example, we found that an apaQTL in progenitors mediated expression of *CEP250* gene (Figure S3B).

We also compared the pLOUEF scores<sup>92</sup> of the genes harboring edSites (edGenes) and APA sites (aGenes) in this study and eGenes and genes for which splice sites (sGenes) were found to be significantly regulated in our previous study<sup>26</sup>. We observed that sGenes, edGenes and aGenes showed lower pLOUEF scores than eGenes in both cell types. Lower pLOUEF scores indicate genes that are generally protected from rare damaging variation, suggesting that they are important for diseases. These findings indicate that the genes affected by editing and APA are likely to be more disease relevant (Figure 5E).

### Interpretation of the function of the brain-relevant GWASs using post-transcriptional QTLs

To explain the function of genetic variants associated with brain-relevant traits, we next leveraged cell-type-specific edQTLs and apaQTLs with brain-related trait GWAS. Applying colocalization analysis to 2,260 brain-trait GWAS including neuropsychiatric disorders, brain structure and function, and cognitive performance (see Methods), we identified 6 and 6 GWAS loci-traits pairs colocalized with progenitor and neuron edQTLs, respectively; also we found 6 and 3 GWAS loci-trait pairs colocalized with progenitor and neurons apaQTLs, respectively. Importantly, we did not detect some of these loci in cell-type-specific eQTL and sQTL analysis<sup>26</sup>, suggesting that our approach to integrate post-transcriptional gene regulatory phenotypes revealed the regulatory mechanism of additional brain-relevant trait GWAS loci (Figure S4A, Table S6). As a specific example for edQTLs colocalized with brain-relevant traits, we observed that a neuron-specific edQTL, rs56320407 was co-localized with an educational attainment GWAS-associated locus (index variant rs2589091, p-value =  $3.3 \times 10^{-8}$ , LD  $r^2 = 0.85$  based on European population) within the *CCDC88A* gene locus<sup>93</sup> (Figure 6A). Importantly, the edQTL was not associated with any significant changes in *CCDC88A* expression showing that edQTL enabled the detection of new brain-related genetic variation (Figure 6A). The edit site (chr2:55406089:A>G, Figure S1B) was within the protein coding sequence of one the isoform of

*CCDC88A*, though did not change its amino acid sequence, but overlapped the intronic region of the rest of the isoforms. The T allele of rs56320407 was associated with an increase in RNA-editing and decreased educational attainment (Figure 6B). Both the edit site and the index edQTL variant rs56320407, which is 20 bp away from the edit site, were within an Alu repeat (Figure 6C). We predicted the secondary structure of a potential IRAlu hairpin using an *in silico* analysis separately for T and C alleles of variant rs56320407 using RNA sequence between this Alu site and the closest Alu repeat in the opposite direction<sup>94</sup> (Figure 6C). The predicted secondary structure of the IRAlu hairpin including T (U in RNA sequence) allele matched with the A nucleotide; whereas structure including the C allele caused a C-A mismatch (Figure 6C). Given that the T allele was associated with higher editing, this observation suggests that the T allele caused formation of an RNA secondary structure substrate which was preferred by the ADAR enzymes that consequently led to higher editing level. *CCDC88A* is an actin binding protein, and it played a role in axonal development and newborn neuron migration during mouse adult neurogenesis<sup>95,96</sup>. Though the edit site did not lead to amino acid change in the protein or differences in mRNA expression, we suggest that the edit site within the *CCDC88A* gene may impact higher cognitive function via altering mRNA stability and eventually translation of protein during cortex development.

One example of an apaQTL colocalized with brain-relevant trait GWAS was found at the *EP300* gene locus. We detected that progenitor-specific index apaQTL variant, rs35508493, was colocalized with variant rs9607782 which is an index SNP within the schizophrenia GWAS<sup>3</sup> ( $p$ -value =  $5.5 \times 10^{-13}$ ) (Figure 6D). We did not observe any genetic variants associated with summarized gene expression at the locus. Insertion of GTA nucleotides at rs35508493 was associated with decrease in usage of the longer 3'UTR isoform and lower risk for schizophrenia (Figure 6E). Variant rs35508493 was overlapped with binding sites of multiple RBPs including LIN28B, YTHDF1, YTHDF2, YTDC1 and IGF2BP3 based on CLIPdb database (Figure 6D)<sup>88</sup>,

suggesting that it may alter APA site usage by interfering with RBP function. *EP300* gene encodes a histone acetyltransferase protein, and its inhibition promoted proliferation of neural progenitors in adult zebrafish<sup>97</sup>. We observed several microRNA binding sites within the genomic location that differ between long and short APA sites (Figure 6F). This observation suggests that different APA site usage may influence mRNA stability by interfering microRNA function which may consequently lead to differences in protein translation that may influence schizophrenia risk by impacting neural proliferation.

We also evaluated how many additional brain-relevant GWAS loci's function could be explained via our cell-type-specific system in addition to adult brain eQTLs. We found that our cell-type-specific QTL approach allowed discovering the function of 0.6-4.5% of GWAS loci which could not be explained by adult brain eQTLs previously (Figure S4B). Furthermore, we also investigated that 1.3%-37.5% of adult brain eQTLs which could already explain the function of brain-relevant trait GWAS loci were also cell-type-specific QTLs, indicating a developmental origin of these adult eQTLs (Figure S4B).

## Discussion

In this study, we identified the impact of genetic variants on cell-type-specific RNA-editing and alternative polyadenylation. We found that: (1) RNA-editing was more frequently observed in neurons compared to progenitor cells. This increase in RNA editing may be mediated by higher expression of *ADAR2* and *ADAR3* in neurons. (2) Consistent with previous findings of 3'UTR isoforms lengthening during differentiation<sup>63-65</sup>, the majority of longer 3'UTR isoforms of the genes were upregulated in neurons. (3) Both edQTLs and apaQTLs were strongly cell-type-specific. (4) Both edQTLs and apaQTLs were enriched within genomic regions and regulatory elements which were different from eQTLs, suggesting independent genetic regulatory mechanisms. (5) We found that a few edQTLs and apaQTLs were



colocalized with brain-relevant trait GWAS loci in progenitor and neuron cells, increasing the known gene regulatory mechanisms underlying complex brain traits.

Previous studies have reported that RNA-editing increases from fetal to adult human brain<sup>41,55</sup>. However, the mechanism causing the developmental increase of RNA-editing has remained unexplored. Our cell-type-specific design using approximately 80 different donors provided sufficient power to observe the difference in RNA-editing between cell types of the prenatal brain. RNA-editing was six times more frequently observed in neurons compared to progenitor cells during cortical development. This observation was consistent with *ADAR2* and *ADAR3* upregulation, but slight *ADAR1* downregulation in neurons. We observed that higher *ADAR1* expression was associated with increased editing levels in progenitors, but not in neurons. Also, we noted that higher *ADAR2* and *ADAR3* expression were associated with higher editing levels, specifically in neurons. Though increased *ADAR2* was also previously found to be associated with increased editing in the adult brain, *ADAR3* has been previously thought to inhibit RNA-editing using expression measured in the adult brain<sup>54,77,78,98</sup>. However, here we show that *ADAR3* is positively correlated with editing in neurons suggesting that *ADAR3* can act as an activator of RNA-editing specifically during early development, but it may later decrease editing levels in adulthood. Future studies measuring RNA-editing levels following modulation of *ADAR3* in immature and mature neurons will provide insights into addressing this controversy. Increased editing levels in neurons suggests that RNA-editing has functionality in neurons including neuronal differentiation, maturation and activity. For instance, we found that higher editing rate within the *CEP104* gene was positively correlated with higher number of neurons generated during differentiation, while the same edit site was not discovered in progenitor cells (Figure 1E). Also, these results imply that the molecular engineering tools such as RADAR and cellREADR<sup>99,100</sup> may be more useful in cell types with higher ADAR expression, such as neurons, and that design of sense-edit-switch sequences that target

endogenous mRNAs may benefit from knowing which variants increase or decrease editing events.

A previous report indicated that longer 3'UTR isoforms are abundant in neurons compared to progenitors<sup>65</sup> that 3'UTRs upregulated in neurons were mainly the longest possible isoform of the genes. These observations suggest that testing 3'UTR isoform levels derived from long-read sequencing across different genotypes will also be a useful strategy to validate apaQTLs given that our short-read RNA-sequencing data may have limited accuracy for quantification of the reads mapping to multiple isoforms. Finally, we detected that apaQTLs showed a higher overlap with eQTLs compared to the overlap of edQTLs with eQTLs. Different microRNA binding sites may exist more likely across different APAs which may eventually impact regulation of gene expression via those microRNAs. On the other hand, only single base changes via RNA-editing might not be sufficient to alter microRNA binding sites whereas they can still influence RNA secondary structure and eventually translation of proteins. Performing ribonucleoprotein immunoprecipitation assays<sup>101</sup> at candidate genetically altered APA and edit sites for the microRNAs will help to explore these potential molecular mechanisms in the future.

Our cell-type-specific approach also increased discovery of genetic regulation on post-transcriptional modulation during human brain development. We observed many more ed/apaQTLs in our cell type specific dataset as compared to fetal bulk brain dataset. This could be due to homogeneous cell populations yielding more accurate quantification of post-transcriptional phenotypes, whereas bulk populations intermingle multiple different cell types each of which has different mechanisms. However, it is also important to note that the lower read depth difference in fetal bulk data compared to the cell-type-specific data might have impacted the number of QTLs discovered. Future studies with greater cellular resolution will likely yield greater discovery of genetically altered post-transcriptional gene regulation.

We observed a very low overlap between edQTLs/apaQTLs and eQTLs, suggesting their impact is independent of the genetic regulation of gene expression as also proposed by several previous reports<sup>22,67–69</sup>. This observation suggests that edQTLs/apaQTLs may impact protein abundance, localization, or function rather than mRNA expression levels. The influence of APA on protein abundance and ribosome occupancy without altering mRNA expression levels has been previously described<sup>67</sup>. Although there is no clear evidence for the impact of RNA-editing on translation yet, changes in protein levels via RNA-editing were detected in a previous study<sup>102</sup>. Previous studies in adult brain data showed the differences in the impact of genetic variants on gene expression and proteins<sup>103–105</sup>. Comparison edQTLs/apaQTLs with protein QTLs in cell-types of developing brain will clarify what the functional consequences of these variants are at the molecular level.

As a result of their independent regulation, novel post-transcriptional QTL colocalizations can be detected with brain-relevant trait GWAS that were not observed in steady state eQTL colocalizations. Importantly, we observed a few edQTLs and apaQTLs colocalized with brain-relevant trait GWAS, contributing to solving the missing regulation problem. Utilization of edQTLs/apaQTLs with larger sample size may enable explanation of additional genetic mechanisms underlying complex brain traits in the future.

### **Acknowledgements**

This work was supported by NIH (R00MH102357, U54EB020403, R01MH118349, R01MH120125). The following core facilities were utilized for this project: UNC Neuroscience Center Microscopy Core (P30NS045892), UNC Mammalian Genotyping Core, CGIBD Advanced Analytics Core (NIH grant P30 DK034987), UNC Flow Cytometry Core Facility, UNC Vector Core, UNC Research Computing. Additional core facilities utilized for this project were: UCLA CFAR (5P30 AI028697), and the UCLA Neuroscience Genomics Core.

## **Author Contributions**

JLS and NA conceived the study. JLS directed and supervised the study. JLS provided funding. DL, BL and JV processed ATAC-seq data. OK and JM performed immunocytochemistry. JM and NA quantified marker expression. MIL aided in editing quantitative loci methodology. NA managed the integration of the datasets, interpreted results, performed genotype imputation, identified RNA-editing and alternative polyadenylation sites, performed edQTL, aQTL, co-localization, mediation, and enrichment analyses. JLS and NA wrote the manuscript. All authors commented on and approved the final version of the manuscript.

## **Declarations of Interest**

The authors declare no competing interests.

## **Availability of data and materials**

Codes are available here [https://bitbucket.org/steinlabunc/post\\_transcriptional\\_qtls/src/master/](https://bitbucket.org/steinlabunc/post_transcriptional_qtls/src/master/)

## **Figure legends**

### **Figure 1. Study design and features of RNA-editing sites**

(A) Study design to identify cell-type-specific edQTLs and apaQTLs.

(B) Proportions of RNA-editing events discovered in each cell type.

(C) Pie chart showing proportions of RNA-editing events within Alu repeats, other repeats or non-repeat regions in the genome per cell type.

(D) Proportions of RNA-editing events across genomic regions. CDS: Coding sequence, 3'UTR: Three prime untranslated region and 5'UTR: Five prime untranslated region

(E) Overlap of RNA-edit sites with GTEx Cortex or BrainVar datasets. The edit sites overlapped are defined as annotated and the sites which do not overlap are defined as novel.

(F) Local motif enrichment for annotated and novel edit sites per cell-type. Number of edit sites (n) in each category is reported.

## **Figure 2. Cell-type-specific RNA-editing during human neurogenesis**

(A) Comparison of Alu editing index (AEI) between progenitors and neurons. T-test p-value was reported.

(B) Differential expression of *ADAR1*, *ADAR2* and *ADAR3* genes between progenitors and neurons. Adjusted p-value (adj.pval) and log fold change (logFC) from limma were reported.

(C) Overlap of RNA-editing sites discovered in progenitors, neurons and fetal bulk data.

(D) Enrichment of cell-type-specific RNA-editing sites within edit sites dysregulated in brain-relevant diseases. The number of edit sites overlap is reported, and the proportion of disease-specific edit sites overlapped with cell-type-specific edit sites is shown on the y-axis. Asterisks indicate significant enrichment (FDR < 0.05). Autism (Autism Spectrum Disorder), Fragile X<sub>1</sub> (Fragile X syndrome from UC Davis database), Fragile X<sub>2</sub> (Fragile X syndrome from NIH biobank dataset), GBM (Glioblastoma), SCZ ACC (Schizophrenia from anterior cingulate cortex), and SCZ DLPFC (Schizophrenia from dorsolateral prefrontal cortex).

(E) Editing rate of the edit site (chr1:3813767:A>G) at the 3'UTR of *CEP104* gene was positively correlated with the proportion of TUJ1+ neurons. Gene model for *CEP104* and the genomic position of the edit site are given at the left. Scatter plot illustrating the correlation between edit rate and the relative abundance of TUJ1+ neurons, and correlation coefficient (r) and p-values are shown. Representative immunocytochemistry images for TUJ1 (in green) and DAPI (in blue) staining of Donors 18 and 321 (D18 and D321) are shown, scale bar is 100 μm.

### Figure 3. Cell-type-specific edQTLs

(A) Overlap of edSites detected in progenitors, neurons, and fetal bulk data.

(B) Primary edQTLs were more significantly associated with editing as they were closer to the edit sites. Association p-values at  $-\log_{10}$  scale are shown on the y-axis and distance from edit sites are shown on the x-axis for progenitors (left) and neurons (right), density of the data points are indicated by color density per cell-type.

(C) Gene ontology results for edGenes found in neurons.

(D) Enrichment of significant edQTLs within different RNA secondary structures illustrated at left. Enrichment p-values at  $-\log_{10}$  scale are shown on the y-axis across structures and data are colored by cell-types. The number of variants (n) significantly associated with edit sites and located within each structure are shown.

### Figure 4. Cell-type-specific alternative polyadenylation and apaQTLs

(A) Principal component analysis for alternative polyadenylation (APA) site usage colored by cell-type.

(B) Differentially expressed genes between progenitors and neurons that encode proteins playing a role in alternative polyadenylation. Fold change (logFC) is given on the x-axis and data points are colored by adjusted p-value at  $-\log_{10}$  scale.  $\log_{10}FC > 0$  indicates the genes upregulated in neurons and  $\log_{10}FC < 0$  indicates the genes upregulated in progenitors, and  $\log_{10}FC = 0$  is shown with dashed vertical line.

(C) Two APA sites of *CALM1* which were differentially expressed between progenitors and neurons are shown. Gene model for *CALM1* is provided above, and relative read count per APA sites for each cell-type are shown where relative read coverage calculated as ratio of number of count supporting each APA site to total number of reads including both APA sites are shown on

the y-axis; and genomic position of the reads are shown on the x-axis. Differential expression of APA<sub>2</sub> (longer 3'UTR isoform) is shown between cell-types.

(D) Overlap of APA sites regulated by primary aSNPs across progenitors, neurons and fetal bulk brain data.

(E) Primary apaQTLs were more significantly associated with APA as they were closer to APA sites. Association p-values at -log<sub>10</sub> scale are shown on the y-axis and distance from APA sites are shown on the x-axis for progenitors (left) and neurons (right), density of the data points are indicated by color density per cell-type.

(F) apaQTL overlapping with canonical polyadenylation signal motif AAUAAA for gene *RPL221* is shown for each cell type. Read coverage per genotype is shown.

### **Figure 5. Comparison of cell-type-specific molecular QTLs**

(A) Distribution of primary eQTLs, sQTLs, edQTLs and apaQTLs from transcription start site (TSS) per cell-type, the distance from TSS is shown on the x-axis.

(B) Distribution of primary eQTLs, sQTLs, edQTLs and apaQTLs from transcription termination site (TTS) per cell-type, the distance from TTS is shown on the x-axis.

(C) Distribution of primary eQTLs, sQTLs, edQTLs and apaQTLs from splice sites per cell-type, the distance from splice is shown on the x-axis. Two distances were calculated relative to intron start and end sites, and the shortest distance was used for comparison for each QTL data.

(D) Overlap of primary e/s/ed/apaQTLs via  $\pi_1$  statistics (progenitors in purple and neurons in green). Matrices are colored based on the proportion of progenitor and neuron primary edSNP-edGene, aSNP-aGene, sSNP-sGene, eSNP-eGene pairs that were non-null associations ( $\pi_1$ ) in each of QTL datasets.

(E) pLOUEF values for aGenes, edGenes, eGenes and sGenes per cell type are shown. P-values from t-test are reported.

**Figure 6. Colocalization of cell-type-specific edQTLs and apaQTLs with brain-relevant trait GWAS loci**

(A) Genomics tracks illustrating that an edQTL within *CCDC88A* gene locus in neurons was colocalized with an index variant for education attainment GWAS; whereas there was not any significant eQTL for *CCDC88A*. Data points were colored based on the pairwise LD  $r^2$  with rs563320407; p-values for each association are shown on the y-axis and genomic positions of genetic variants are shown on the x-axis.

(B) Boxplot illustrating distribution of editing rate across rs56320407.

(C) The predicted secondary structure of the IRAlu hairpin for T and C alleles of rs56320407. Edit sites and genetic variants are indicated by red and blue colors, respectively.

(D) Genomics tracks illustrating that an apaQTL within *EP300* gene locus in progenitor was colocalized with an index variant for schizophrenia GWAS; whereas there was not any significant eQTL for *EP300* gene. Data points were colored based on the pairwise LD  $r^2$  with rs35508493; p-values for each association are shown on the y-axis and genomic positions of genetic variants are shown on the x-axis.

(E) Boxplot illustrating distribution of editing rate across rs56320407.

(F) microRNA binding sites are shown for the genomic region which differs between two potential 3'UTR isoforms of *EP300* gene (genomic coordinates for APA site 1: chr22:41,178,956-41,180,079 and APA site 2: chr22:41,178,956-41,179,495).



**Figure S1. Validation of RNA-editing sites discovered in RNA-seq data via ATAC-seq data.**

(A) The number of donors supporting that genomic DNA includes unedited alleles are shown on the x-axis. The percentage of unedited allele per edit sites is shown on the right y-axis (red) and the total number of ATAC-seq supporting unedited allele is given on the left y-axis (blue).

(B) Example ATAC-seq reads for RNA-editing sites chr19:18365401:A>G, chr6:163422494:A>G and chr2:55406089:A>G.

**Figure S2. ALU editing index is correlated with ADAR gene expression and developmental time.**

(A) Correlation of AEI values with VST normalized ADAR1, ADAR2 and ADAR3 expressions in progenitors (purple) and neurons (green). Correlation coefficients ( $r$ ) and p-values are shown per relationship.

(B) Alu editing index across different gestational weeks in fetal bulk brain data.

(C) Correlation of technical confounders with *ADAR1-3* gene expressions per cell-type. Correlation efficient ( $r$ ) and p-values are reported.

**Figure S3. Comparison of cell-type-specific molecular QTLs**

(A) Overlap of primary eQTLs, sQTLs, edQTLs and apaQTLs with chromatin accessibility regions are RBP binding sites per cell type. Chi-square test p-values ( $p$ ) are reported for each pairwise comparison.

(B) APA site (chr20:35511626-35519280) within the *CEP250* gene mediated its expression. Genomic tracks for progenitor eQTLs and apaQTL are shown, and tracks are colored based on relative LD  $r^2$  for the variant rs2236160, which is the nearest variant to APA site.

(C) Posterior probability of forward, independence and reactive models for mediation of *CEP250* gene expression via APA site. Lower cartoon illustrates the relationship between variant (X) within the APA site, mediator APA site (M) and *CEP250* gene.

### **Figure S4. Comparison of GWAS colocalizations across cell-type-specific molecular QTLs**

(A) Number of brain-relevant trait-loci pairs are found in eQTLs, sQTLs, edQTLs and apaQTLs, and shared pairs with eQTLs are indicated by black color, and pairs which were detected by using QTLs other than that eQTLs are indicated by blue color.

(B) Number of brain-relevant trait-loci pairs are found in cell-type-specific eQTLs, sQTLs, edQTLs and apaQTLs in addition to loci explained by adult brain eQTLs.

### **Methods**

#### Preparation of primary human neural progenitor cells (phNPCs)

We established phNPCs culture including neural progenitor cells and neuronal progeny differentiated from these progenitors by following the experimental workflow described in our previous work<sup>26,70,106</sup>. We acquired the human fetal brain tissue (14-21 gestation weeks old) derived from voluntary terminated pregnancy according to the IRB regulations at UCLA through the Gene and Cell Therapy Core. The tissue pieces corresponding to the cortex were visually selected for generation of phNPCs. In the Geschwind lab at UCLA, we dissociated these tissues and formed neurospheres by using them as we have previously described<sup>26,106</sup>. We then plated the neurospheres on the plates coated with laminin/fibronectin and polyornithine, and after an average of 2.5 +/- 1.8 standard deviation passages, they were cryopreserved to transfer to Stein lab at UNC-Chapel Hill<sup>26</sup>.

phNPCs from 89 unique donors were further randomly grouped into 8-9 donors for 12 rounds, and each round was thawed every 3 weeks. Each round was processed by mostly the

same person and on the same day of the week, where changes in these technical variable changes were documented. We cultured progenitor cells in proliferation media as previously described<sup>26,70</sup> for three weeks, and then prepared RNA-seq libraries. To differentiate progenitor cells into neurons, we first cultured the cells in the media without growth factors for 5 weeks and then transduced cells with AAV2-hSyn1-eGFP virus (with 20,000 multiplicity of infection (MOI)) that carries a reporter gene expressed specifically in neurons<sup>26</sup>. After viral transduction, we differentiated the cells for 3 weeks longer, and isolated EGFP-labeled neurons using FACS sorting machines BD FACS Aria II or Sony SH800S. We kept the neurons within the Qiazol solution to prepare RNA-seq libraries.

#### RNA-sequencing and data processing

We prepared RNA-seq libraries and sequenced them as described in our previous work<sup>26</sup>. We obtained 150 bp paired end reads with a mean read depth of  $99.8M \pm 29.8$  SD read pairs per library.

To process the RNA-sequencing data, as described in our previous work<sup>26</sup>, if the same library was sequenced on multiple flow cells, we merged .fastq files, trimmed the adapters for all libraries using Cutadapt/1.15 software<sup>107</sup>, and performed quality control with FastQC software. We aligned the RNA-seq reads into GRCh38 release92 reference genome including sequence of AAV2-hSyn1-eGFP plasmid using STAR/2.6.0a aligner program<sup>108</sup>.

To assess consistency of genotypes detected by genotyping array and RNA-seq, we performed VerifyBamID analysis (v.1.1.3)<sup>109</sup>, as described previously<sup>26</sup>. We retained the RNA-seq libraries with [CHIPMIX] < 0.04 and [FREEMIX] < 0.04, and assigned correct donor IDs for 8 libraries where there was a sample swap. Additionally, one library was missing cDNA concentration and removed, also libraries with eRIN score lower than 7 were not included. Following quality control, we obtained 84 and 74 unique donors for progenitors and neurons,

respectively. We applied the same workflow to process RNA-seq data derived from fetal bulk brain samples and retained 235 unique donors as previously described<sup>26</sup>.

### Genotyping and imputation

We performed genotyping by utilizing HumanOmni2.5Exome or Illumina HumanOmni2.5 platforms followed by filtering via PLINK v.1.90b3 software<sup>110</sup> with the parameters `–hwe 1 x 10-6 –geno 0.05 –mind 0.01 –maf 0.01` as described previously<sup>26,70</sup>. We utilized the TOPMed imputation server for imputation with the TOPMed reference panel (Version R2 on GRC38)<sup>111</sup>, after processing the genotype data via imputation preparation pipeline by using the algorithm `perl HRC-1000G-check-bim_v3.pl -b <bim file> -f <frequency file> -r 1000GP_Phase3_combined.legend.gz -g -p ALL` (<https://www.well.ox.ac.uk/~wrayner/tools/>). For downstream analyses, we retained the variants with following criteria: minor allele frequency (MAF) > 0.01, imputation quality score  $R^2 > 0.3$  and Hardy-Weinberg equilibrium at  $p > 1 \times 10^{-6}$ .

### Detection and quantification of RNA-editing events by using RNA-sequencing data

To detect RNA-editing events by using RNA-sequencing data, we first aimed to reduce mapping bias using the WASP method<sup>112</sup> and remapped RNA-seq reads mapped via STAR/2.6.0a aligner<sup>108</sup>. Following remapping, we discarded duplicated reads and extracted uniquely mapped reads as input for Reditools software v2.0<sup>71</sup>. Using Reditools software, we identified and quantified RNA-editing sites by applying the following parameters: `-S -s 2 -q 20 -bp 25 -ss 5 -mrl 50 -C -T 2 –os 5` where `-S` was used for including only edit sites in the output, `-s` was for strand inference, `-q` was for minimum read quality score, `-bp` was for minimum base quality score, `-ss` was for splicing span, `-mrl` was for minimum read length, `-C` was for strand correction, `-T` was for strand confidence and `–os` was for omopolymeric-span. We used Homo sapiens gene ensembl v.92 as the reference genome. We discarded multi-allelic sites and the sites overlapped with genomic variants from our genotype data with imputation  $R^2$  greater than 0.3 and common SNPs in dbSNP (v153) unless they were listed in the REDportal database<sup>74</sup>.

For downstream analysis, we retained RNA-editing sites supported by at least 85% of donors with 10 RNA-seq counts (at least 2 counts for edited allele) per cell-type and fetal bulk tissue. We defined RNA-editing rate as the ratio of the number of read counts supporting the edited allele to the sum of the number of the read counts supporting both edited and unedited alleles.

#### Validation of RNA-edit sites via matched ATAC-seq data

To confirm that RNA-edit sites are nucleotide changes in RNA-sequence but not genomic mutations, we extracted WASP-mapped ATAC-seq reads from our two previous studies<sup>70,72</sup> which overlapped with per edit site. We calculated the abundance of each unedited allele of each edit site per donor by getting the ratio of number of reads supporting the unedited allele to total coverage at that site (Figure S1A). We excluded the nucleotides with base quality lower than 30, and the reads where any mismatch were reported in CIGAR string since they obscure the finding of the nucleotides at the genomic region of interest.

#### Calculation of the ALU editing index

We computed Alu editing index (AEI) per RNA-sequencing sample bam file per cell-type and fetal bulk tissue (including uniquely mapped reads after STAR alignment and WASP algorithms as used for the detection of RNA-editing via Reditools) via the RNAEditingIndexer algorithm<sup>76</sup>.

#### Enrichment of local motif for RNA-edit sites

For each edit site, we extracted RNA-sequence within +/- 4 bp window of the edit sites. Providing these sequences to the EDLogo software, we quantified and visualized local sequence motifs<sup>113</sup>.

#### Enrichment of progenitor and neuron RNA-edit sites within disease-relevant edit sites

To assess the significance of overlap between disease-relevant edit sites (defined based on adjusted p-value < 0.05 in differential editing analysis between case and control) and edit sites detected in each cell type, we applied GeneOverlap R package.

### Detection of alternative polyadenylation sites from RNA-seq data for each cell type

To detect and quantify alternative polyadenylation by using RNA-seq data, we followed the pipeline described by QAPA method<sup>65</sup>. We initially built 3'UTR libraries including potential alternative polyadenylation sites for each gene by using biomaRt ensembl gene metadata table (human version 92), GTF file Homo\_sapiens.GRCh38.92 as gene prediction table, and PolyASite database (hg38) and GENCODE poly(A) sites track (hg38) as poly(A) site annotations followed by extraction of 3'UTR sequences by integrating these annotations with reference genome fasta file (GRCh38 release92). After trimming the sequencing adaptors from RNA-seq libraries via Cutadapt/4.1<sup>107</sup>, we quantified 3'UTR isoforms via salmon/1.9.0<sup>114</sup> based on the 3'UTR library generated by QAPA by correcting for GC and sequence-specific biases. To infer poly(A) usage (PAU) value for a given 3'UTR isoform of a gene, we divided the expression of this 3'UTR isoform by the sum of the expression of all other 3'UTR isoforms detected for that gene via *qapa quant* function<sup>65</sup>. For the downstream analyses, we applied the following steps: (1) we retained the 3'UTR isoforms supported by 10 counts at least 10% of the donors for each cell-type. (2) For the genes in which we detected only two 3'UTR isoforms, we randomly selected one of the isoforms to prevent statistical bias since these two isoforms were complementary to each other. (3) We normalized the PAU ratios for the remaining 3'UTR isoforms via quantile normalization.

### Differential alternative 3'UTR expression analysis

Prior to the differential alternative 3'UTR expression analysis, we first corrected quantile normalized neuron gene expression data for the batch effect caused by the usage of different machines for FACS sorting via *removeBatchEffect* function of limma R package<sup>115</sup>. Followed by batch correction for neurons, we combined two cell-type-specific data, and performed a paired differential gene expression via limma package by using the design matrix `model.matrix(~Cell-type + as.factor(DonorID) + RIN, dataset)`. We identified differentially expressed 3'UTR isoforms

between cell-types as adjusted p-values  $< 0.05$  after multiple test correction via Benjamini-Hochberg method<sup>116</sup>. We assessed the isoform lengthening or shortening based on if the differentially expressed 3'UTR had the longest or the shortest 3'UTR isoform among all potential 3'UTR isoforms for a gene.

#### Immunohistochemistry of neurons for TUJ1 marker and quantification

We used the same experimental procedure to generate neurons from phNPCs as described previously<sup>26,70</sup>. At 8 weeks of differentiation, we fixed neuron cells in 4% PFA, and permeabilized them by using 0.4% Triton in PBST solution and performed blocking within 10% goat serum dissolved in PBST. After we incubated primary antibodies for TUJ1 (1:2000, Catalog # 801202) overnight in 3% goat serum dissolved in PBST solution at 4°C, we washed the cells three times with PBST. We applied fluorophore-conjugated secondary antibodies (Alexa Fluor 488, goat anti-mouse, 1:1000, Invitrogen, Catalog # A11001) at room temperature for an hour, and applied DAPI staining for 10 minutes.

We performed imaging by using Nikon Eclipse Ti2 with pco.edge 4.2Q High QE sCMOS camera via 10x objective. Prior to segmentation, we used ImageJ to isolate the DAPI channel, transformed them to grayscale and divided images into 4 crops at 2862 by 2862 pixels. We applied Cellpose software<sup>117</sup> for segmentation by implementing a nuclear segmentation method in which we set the nucleus diameter as 9 microns. We subtracted the cell outlines from the generated cell masks resulting in a final nuclear mask. To count TUJ1+ cells, we applied CellProfiler to masks generated by Cellpose. We excluded cells if they had nuclei smaller than 6 microns in diameter, which were likely dead cells. In the other image channels, objects with high intensity were considered debris and masked out of the images to aid in threshold and background intensity calculations. For each channel, we corrected images for illumination inhomogeneity, measured background intensity, and images intensities. We classified each cell

for the TUJ1 marker if the average intensities were at least 1.5 standard deviations above the median of the background intensity.

### Cell-type-specific editing quantitative loci analysis

We tested association of editing rate with genetic variants within +/- 100 kb window of editing site and located in the same gene harboring edit site to perform editing QTL (edQTL) analysis. We included the editing sites if at least 85% of samples had at least 10 read counts (at least 2 counts to support the edited allele) to support the editing site for each cell type. We retained only the variants if at least two heterozygous donors and at least two homozygous minor allele donors, or no homozygous minor allele donors were present as a filtering strategy we previously used<sup>26</sup>. Since the donors showing the sufficient read counts might be different across editing sites, for each editing site, we used 85% of the donors that every one of them supported editing site with at least 10 counts in each cell type (at least 2 counts for edited allele).

We performed cell-type-specific edQTL mapping analysis by using a generalized linear model with binomial distribution that controls for population stratification and unmeasured technical variation. To control for population stratification, we calculated MDS of global genotype and used the first three MDS components as covariates. We controlled the unmeasured technical variation which affects RNA-editing via an optimization strategy. For each cell-type, we utilized principal component analysis (PCA) for unmeasured technical variation, and computed global editing PCs via `prcomp()` function from stats R package by using edit rate values per edit site. During the optimization strategy, we re-performed edQTL analysis by sequentially adding the global editing PCs, first 3 MDSs of global genotype for each cell-type. We detected FACS sorter as a major technical factor impacting editing rate in neurons, and controlled for it for neuron edQTL analysis (p-value =  $1.8 \times 10^{-7}$  for PC1 of global editing and FACS sorter relationship). After each run, we calculated the number of edit sites significantly associated with



at least one genetic variant (edSite) at a 5% false discovery rate. Since we found that including 1 PC of global editing maximized the edSite discovery in both progenitor and neurons:

The optimized model we used for progenitors was:

Edit rate  $\sim$ SNP + 3 MDS of global genotype + 1 PC of global editing

The optimized model we used for neurons was:

Edit rate  $\sim$ SNP + 3 MDS of global genotype + FACS sorter + 1 PC of global editing

As implemented in our previous work<sup>26</sup>, we applied a hierarchical correction procedure termed eigenMT-FDR<sup>118</sup>, which allowed us to stringently control for multiple comparisons by considering both the number of edit sites and the variants tested. In this algorithm, we first computed locally adjusted p-values for cis-SNPs per edit site via the eigenMT approach in which a genotype correlation matrix was used to estimate the effective number of independent tests<sup>119</sup>. Then, we performed FDR procedure by using locally adjusted p-values that resulted in globally adjusted p-values per edit site. As the last step, the edit sites with a globally adjusted p-value lower than 0.05 were defined as edSites. We conducted the same procedure to discover edQTLs in fetal bulk brain data.

#### Cell-type-specific alternative polyadenylation quantitative loci analysis

To perform cell-type-specific alternative polyadenylation QTL (apaQTL) analysis, we tested association of quantile normalized PAU values per 3'UTR isoform with the genetic variants within the +/-25 kb window of isoform start and end sites. We retained the genetic variants if they were carried by at least two heterozygous donors and without any homozygous minor allele donors, or if they were carried by at least two minor allele homozygous donors identical to edQTL analysis.

We performed apaQTL mapping analysis by controlling for population stratification and cryptic relatedness via a linear mixed effects regression model by using EMMAX software<sup>120</sup>. We controlled the population stratification by using the first three MDS components of global genotype as covariates. We generated the identity by state (IBS) kinship matrix by implementing `emmax-kin -v -h -d` algorithm by using variants from the non-imputed genotype data, and excluded the variants on the same chromosome via MLMe method<sup>121</sup>. Similar to edQTL analysis, we performed an optimization strategy to identify the number of PCs of global expression of 3'UTR isoforms, which was computed via `prcomp()` function of stats R package by using quantile normalized PAU values per edit site separately for each cell type. After sequentially adding PCs of global expression of 3'UTR isoforms as covariates to re-run apaQTL analysis, we identified 9 PCs and 6 PCs of global 3'UTR isoform expression in progenitor and neurons showed the highest number of APA site significant associated with at least one genetic variant at 5% FDR. As a result, we applied the following models per cell-type:

The optimized model we used for progenitors was:

$$\text{PAU} \sim \text{SNP} + 3 \text{ MDS of global genotype} + 9 \text{ PC of global 3'UTR expression} + \varepsilon$$

The optimized model we used for neurons was:

$$\text{PAU} \sim \text{SNP} + 3 \text{ MDS of global genotype} + 6 \text{ PC of global 3'UTR expression} + \text{FACS sorter} + \varepsilon$$

We conducted same pipeline to discover apaQTLs in fetal bulk data and the optimized model we used for fetal bulk was:

$$\text{PAU} \sim \text{SNP} + 3 \text{ MDS of global genotype} + 5 \text{ PC of global 3'UTR expression} + \varepsilon$$

where PAU is Poly(A) usage<sup>65</sup> and we defined an error term  $\varepsilon$  for which  $\text{cov}(\varepsilon) = \sigma_u^2 u_K + \sigma_e^2 I$  in which kinship matrix used to account genetic relatedness is indicated by  $u_K$ , variance is indicated by  $\sigma_u^2$  and  $\sigma_e^2$  is random noise of the variance.

### Assessment of QTL sharing between cell-types and different molecular QTLs

For cell-to-cell comparison, the proportion of progenitor and neuron primary SNP-edit site pairs or APA site pairs that are non-null associations in neuron and progenitor edQTLs or apaQTLs data was estimated by utilizing the corresponding p-values to SNP-edit site pairs or APA via  $\pi_1$  statistics<sup>89</sup> by using the qvalue R package<sup>122</sup>. Similarly, for edQTL/apaQTL to eQTL comparison, we estimated the proportion of progenitor and neuron primary SNP-edGene pairs, gene including edit site or SNP-aGene pairs, gene including APA site pairs that are non-null associations in progenitor and neuron eQTL data ( $\pi_1$ ) by using the corresponding p-values to SNP-Gene pairs detected in both datasets.

### Prediction of inverted repeat Alus (*IRAlu*) RNA hairpin secondary structures

To predict IRAlu RNA hairpin secondary structure, we extracted RNA-sequences between two Alu repeats in opposite directions, and generated two sequences corresponding to different alleles of a given genetic variant. These sequences were provided to viennaRNA RNAfold software, and secondary structures were predicted and visualized via graphical output from the software<sup>94</sup>.

### Enrichment of edQTLs within RNA secondary structures

As described in previous study<sup>22</sup>, for each edSite (edit site significantly associated with at least one genetic variant), we extracted the RNA sequences including genetic variants within +/- 800 bp window of the edSites. We included only sequences within gene start and end coordinates within this window; therefore, some sequences were shorter than 1601 bp. We matched the alleles for variants within this genomic window if the LD  $r^2$  between them was greater than 0.8. We first converted sequences files to bpseq via contrafold software<sup>123</sup>. Next, providing these bpseq file formats for bpRNA software<sup>124</sup>, we predicted RNA secondary structures including bulge, hairpin loop, interior loop, multiloop, stem within these RNA-sequences. To assess enrichment of significant edQTLs within RNA-secondary structures, for

each structure, we randomly selected non-significant eQTLs in equal number of significant edQTLs in each structure category, which were matching minor allele frequency (MAF) and distance from edit sites with 50% standard error of both features for 1,000 times. We computed the enrichment p-value for each RNA structure as a number of observations in which the overlap of nonsignificant edQTLs with the RNA structure was higher than the overlap of significant edQTLs with the RNA structure divided by 1,000.

### Comparison of molecular QTLs

We extracted cell-type-specific primary eQTLs and sQTL from our previous study<sup>26</sup>, and cell-type-specific primary edQTLs and apaQTLs discovered in the current study for comparison. For distance from TSS/TTS, we calculated the distance between genetic variant and TSS/TTS of genes for which expression was tested in eQTLs, the distance between genetic variant and TSS/TTS of genes where alternative splicing tested was located in sQTLs, the distance between genetic variant and TSS/TTS of genes where edit site tested was located in edQTLs, and the distance between genetic variant and TSS/TTS of genes where APA site tested was located in apaQTLs considering the expression of the gene in either forward or reverse strand. For distance from splice sites, we computed the distance of genetic variants from either intron start and end sites for all potential alternative splicing events for a given gene, and used the shortest distance for comparison. To compare enrichment of primary QTLs within chromatin accessibility sites, we assessed the overlap of genetic variants within chromatin accessibility regions which were differentially accessible in progenitors/neuron<sup>70</sup> for progenitor and neuron QTLs, respectively. To compare enrichment of primary QTLs within RBP binding sites, we utilized CLIPdb data, and assessed the overlap of genetic variants with this dataset<sup>88</sup>. Pairwise comparisons were performed via chi-square test.

### GWAS colocalization analysis

We applied LD-thresholded colocalization analysis to find edQTLs and apaQTLs colocalized with the traits for each cell type separately<sup>26,125</sup>. Summary statistics from GWAS for schizophrenia (SCZ)<sup>3</sup>, educational attainment (EA)<sup>93</sup>, major depression disorder (MDD)<sup>126</sup>, cortical thickness and surface area from UKBB<sup>6</sup>, and the ENIGMA project<sup>5</sup>, neuroticism<sup>127</sup>, IQ<sup>4</sup>, cognitive performance (CP)<sup>93</sup>, bipolar disorder (BP)<sup>1</sup>, attention-deficit/hyperactivity disorder (ADHD)<sup>128</sup>, Parkinson's disease (PD)<sup>129</sup> and Alzheimer's disease (AD)<sup>130</sup> were used. Index GWAS SNPs were defined as two LD-independent genome-wide significant GWAS signals ( $p$ -value  $< 5 \times 10^{-8}$ ) with pairwise LD  $r^2 < 0.2$  calculated by using European population of 1000 Genomes (phase 3). For comparison of colocalizations with eQTLs and sQTLs<sup>26</sup>, we used LD  $r^2 < 0.5$  to detect index GWAS variants. To perform colocalization analysis, first, we detected two variants (one from index variants of GWAS and one from index variants of the QTL study) which had pairwise LD  $r^2$  greater than 0.8 based on either European population or our study). Then, we re-performed the edQTL/apaQTL analysis by conditioning on GWAS index variant, and if the association between edit rate/APA usage and the QTL index variant was no longer significant, we considered these two loci as co-localized.

## REFERENCES

1. Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* *51*, 793–803.
2. Howard, D.M., Adams, M.J., Clarke, T.-K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaaars, S.P., Ward, J., Wigmore, E.M., et al. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* *22*, 343–352.
3. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* *50*, 381–389.
4. Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* *50*, 912–919.
5. Grasby, K.L., Jahanshad, N., Painter, J.N., Colodro-Conde, L., Bralten, J., Hibar, D.P., Lind, P.A., Pizzagalli, F., Ching, C.R.K., McMahon, M.A.B., et al. (2020). The genetic architecture of the human cerebral cortex. *Science* *367*. [10.1126/science.aay6690](https://doi.org/10.1126/science.aay6690).
6. Smith, S.M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K., and Elliott, L.T. (2021). An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature Neuroscience* *24*, 737–745. [10.1038/s41593-021-00826-4](https://doi.org/10.1038/s41593-021-00826-4).
7. Albert, F.W., and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* *16*, 197–212.
8. Cano-Gamez, E., and Trynka, G. (2020). From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* *11*, 424.
9. Sullivan, P.F., and Geschwind, D.H. (2019). Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* *177*, 162–183.
10. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* *19*, 1442–1453.
11. Sng, L.M.F., Thomson, P.C., and Trabzuni, D. (2019). Genome-wide human brain eQTLs: In-depth analysis and insights using the UKBEC dataset. *Sci. Rep.* *9*, 19201.
12. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. (2017). Genetic effects on gene expression across human tissues. *Nature* *550*, 204–213.

13. Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science* 362. 10.1126/science.aat8464.
14. Zeng, B., Bendl, J., Kosoy, R., Fullard, J.F., Hoffman, G.E., and Roussos, P. (2022). Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat. Genet.* 10.1038/s41588-021-00987-9.
15. de Klein, N., Tsai, E.A., Vochteloo, M., Baird, D., Huang, Y., Chen, C.-Y., van Dam, S., Oelen, R., Deelen, P., Bakker, O.B., et al. (2023). Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat. Genet.* 55, 377–388.
16. Connally, N.J., Nazeen, S., Lee, D., Shi, H., Stamatoyannopoulos, J., Chun, S., Cotsapas, C., Cassa, C.A., and Sunyaev, S.R. (2022). The missing link between genetic association and regulatory function. *Elife* 11. 10.7554/eLife.74970.
17. Umans, B.D., Battle, A., and Gilad, Y. (2020). Where Are the Disease-Associated eQTLs? *Trends Genet.* 10.1016/j.tig.2020.08.009.
18. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., de la Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2020). Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* 181, 745. 10.1016/j.cell.2020.04.016.
19. Wen, C., Margolis, M., Dai, R., Zhang, P., Przytycki, P.F., Vo, D.D., Bhattacharya, A., Kim, M., Matoba, N., Tsai, E., et al. (2023). Cross-ancestry, cell-type-informed atlas of gene, isoform, and splicing regulation in the developing human brain. *medRxiv*. 10.1101/2023.03.03.23286706.
20. Bryois, J., Calini, D., Macnair, W., Foo, L., Urich, E., Ortmann, W., Iglesias, V.A., Selvaraj, S., Nutma, E., Marzin, M., et al. (2022). Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat. Neurosci.* 25, 1104–1112.
21. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2022). Limited overlap of eQTLs and GWAS hits due to systematic differences in discovery. *bioRxiv*. 10.1101/2022.05.07.491045.
22. Li, Q., Gludemans, M.J., Geisinger, J.M., Fan, B., Aguet, F., Sun, T., Ramaswami, G., Li, Y.I., Ma, J.-B., Pritchard, J.K., et al. (2022). RNA editing underlies genetic risk of common inflammatory diseases. *Nature* 608, 569–577.
23. Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. *Science* 352, 600–604.
24. Zhang, Y., Yang, H.T., Kadash-Edmondson, K., Pan, Y., Pan, Z., Davidson, B.L., and Xing, Y. (2020). Regional variation of splicing QTLs in human brain. *Am. J. Hum. Genet.* 107, 196–210.
25. Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Castel,



- S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., et al. (2020). Cell type-specific genetic regulation of gene expression across human tissues. *Science* 369. 10.1126/science.aaz8528.
26. Aygün, N., Elwell, A.L., Liang, D., Lafferty, M.J., Cheek, K.E., Courtney, K.P., Mory, J., Hadden-Ford, E., Krupa, O., de la Torre-Ubieta, L., et al. (2021). Brain-trait-associated variants impact cell-type-specific gene regulation during neurogenesis. *Am. J. Hum. Genet.* 108, 1647–1668.
  27. Sommer, B., Köhler, M., Sprengel, R., and Seeburg, P.H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* 67, 11–19.
  28. Shimokawa, T., Rahman, M.F.-U., Tostar, U., Sonkoly, E., Stähle, M., Pivarcsi, A., Palaniswamy, R., and Zaphiropoulos, P.G. (2013). RNA editing of the GLI1 transcription factor modulates the output of Hedgehog signaling. *RNA Biol.* 10, 321–333.
  29. Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. *Cell* 123, 249–263.
  30. Chen, L.-L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* 35, 467–478.
  31. Irimia, M., Denuc, A., Ferran, J.L., Pernaute, B., Puellas, L., Roy, S.W., Garcia-Fernández, J., and Marfany, G. (2012). Evolutionarily conserved A-to-I editing increases protein stability of the alternative splicing factor Nova1. *RNA Biol.* 9, 12–21.
  32. Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. *Nature* 399, 75–80.
  33. Amin, E.M., Liu, Y., Deng, S., Tan, K.S., Chudgar, N., Mayo, M.W., Sanchez-Vega, F., Adusumilli, P.S., Schultz, N., and Jones, D.R. (2017). The RNA-editing enzyme ADAR promotes lung adenocarcinoma migration and invasion by stabilizing FAK. *Sci. Signal.* 10, eaah3941.
  34. Salameh, A., Lee, A.K., Cardó-Vila, M., Nunes, D.N., Efstathiou, E., Staquicini, F.I., Dobroff, A.S., Marchiò, S., Navone, N.M., Hosoya, H., et al. (2015). PRUNE2 is a human prostate cancer suppressor regulated by the intronic long noncoding RNA PCA3. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8403–8408.
  35. Park, E., Jiang, Y., Hao, L., Hui, J., and Xing, Y. (2021). Genetic variation and microRNA targeting of A-to-I RNA editing fine tune human tissue transcriptomes. *Genome Biol.* 22, 77.
  36. Zhang, L., Yang, C.-S., Varelas, X., and Monti, S. (2016). Altered RNA editing in 3' UTR perturbs microRNA-mediated regulation of oncogenes and tumor-suppressors. *Sci. Rep.* 6, 23226.
  37. Nishikura, K. (2016). A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* 17, 83–96.
  38. Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., Isaacs, F.J.,



- Rechavi, G., Li, J.B., Eisenberg, E., et al. (2014). A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* *24*, 365–376.
39. Chen, L.-L., DeCerbo, J.N., and Carmichael, G.G. (2008). Alu element-mediated gene silencing. *EMBO J.* *27*, 1694–1705.
  40. Park, E., Guo, J., Shen, S., Demirdjian, L., Wu, Y.N., Lin, L., and Xing, Y. (2017). Population and allelic variation of A-to-I RNA editing in human transcriptomes. *Genome Biol.* *18*, 143.
  41. Hwang, T., Park, C.-K., Leung, A.K.L., Gao, Y., Hyde, T.M., Kleinman, J.E., Rajpurohit, A., Tao, R., Shin, J.H., and Weinberger, D.R. (2016). Dynamic regulation of RNA editing in human brain development and disease. *Nat. Neurosci.* *19*, 1093–1099.
  42. Behm, M., and Öhman, M. (2016). RNA editing: A contributor to neuronal dynamics in the mammalian brain. *Trends Genet.* *32*, 165–175.
  43. Rosenthal, J.J.C., and Seeburg, P.H. (2012). A-to-I RNA editing: effects on proteins key to neural excitability. *Neuron* *74*, 432–439.
  44. Herrick-Davis, K., Grinde, E., and Niswender, C.M. (1999). Serotonin 5-HT<sub>2C</sub> receptor RNA editing alters receptor basal activity: implications for serotonergic signal transduction. *J. Neurochem.* *73*, 1711–1717.
  45. Marion, S., Weiner, D.M., and Caron, M.G. (2004). RNA editing induces variation in desensitization and trafficking of 5-hydroxytryptamine 2c receptor isoforms. *J. Biol. Chem.* *279*, 2945–2954.
  46. Breen, M.S., Dobbyn, A., Li, Q., Roussos, P., Hoffman, G.E., Stahl, E., Chess, A., Sklar, P., Li, J.B., Devlin, B., et al. (2019). Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nat. Neurosci.* *22*, 1402–1412.
  47. Dracheva, S., Elhakem, S.L., Marcus, S.M., Siever, L.J., McGurk, S.R., and Haroutunian, V. (2003). RNA editing and alternative splicing of human serotonin 2C receptor in schizophrenia. *J. Neurochem.* *87*, 1402–1412.
  48. Eran, A., Li, J.B., Vatalaro, K., McCarthy, J., Rahimov, F., Collins, C., Markianos, K., Margulies, D.M., Brown, E.N., Calvo, S.E., et al. (2013). Comparative RNA editing in autistic and neurotypical cerebella. *Mol. Psychiatry* *18*, 1041–1048.
  49. Tran, S.S., Jun, H.-I., Bahn, J.H., Azghadi, A., Ramaswami, G., Van Nostrand, E.L., Nguyen, T.B., Hsiao, Y.-H.E., Lee, C., Pratt, G.A., et al. (2019). Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* *22*, 25–36.
  50. Lyddon, R., Dwork, A.J., Keddache, M., Siever, L.J., and Dracheva, S. (2013). Serotonin 2c receptor RNA editing in major depression and suicide. *World J. Biol. Psychiatry* *14*, 590–601.
  51. Weissmann, D., van der Laan, S., Underwood, M.D., Salvétat, N., Cavarec, L., Vincent, L., Molina, F., Mann, J.J., Arango, V., and Pujol, J.F. (2016). Region-specific alterations of A-to-I RNA editing of serotonin 2c receptor in the cortex of suicides with major depression.

Transl. Psychiatry 6, e878–e878.

52. Srivastava, P.K., Bagnati, M., Delahaye-Duriez, A., Ko, J.-H., Rotival, M., Langley, S.R., Shkura, K., Mazzuferi, M., Danis, B., van Eyll, J., et al. (2017). Genome-wide analysis of differential RNA editing in epilepsy. *Genome Res.* 27, 440–450.
53. Gaisler-Salomon, I., Kravitz, E., Feiler, Y., Safran, M., Biegon, A., Amariglio, N., and Rechavi, G. (2014). Hippocampus-specific deficiency in RNA editing of GluA2 in Alzheimer's disease. *Neurobiol. Aging* 35, 1785–1791.
54. Cuddleston, W.H., Li, J., Fan, X., Kozenkov, A., Lalli, M., Khaliq, S., Dracheva, S., Mukamel, E.A., and Breen, M.S. (2022). Cellular and genetic drivers of RNA editing variation in the human brain. *Nat. Commun.* 13, 2997.
55. Cuddleston, W.H., Fan, X., Sloofman, L., Liang, L., Mossotto, E., Moore, K., Zipkowitz, S., Wang, M., Zhang, B., Wang, J., et al. (2022). Spatiotemporal and genetic regulation of A-to-I editing throughout human brain development. *Cell Rep.* 41, 111585.
56. Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30.
57. Mayr, C. (2017). Regulation by 3'-untranslated regions. *Annu. Rev. Genet.* 51, 171–194.
58. Melamed, Z. 'ev, López-Erauskin, J., Baughn, M.W., Zhang, O., Drenner, K., Sun, Y., Freyermuth, F., McMahon, M.A., Beccari, M.S., Artates, J.W., et al. (2019). Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. *Nat. Neurosci.* 22, 180–190.
59. Rhinn, H., Qiang, L., Yamashita, T., Rhee, D., Zolin, A., Vanti, W., and Abeliovich, A. (2012). Alternative  $\alpha$ -synuclein transcript usage as a convergent mechanism in Parkinson's disease pathology. *Nat. Commun.* 3, 1084.
60. Romo, L., Ashar-Patel, A., Pfister, E., and Aronin, N. (2017). Alterations in mRNA 3' UTR isoform abundance accompany gene expression changes in human Huntington's disease brains. *Cell Rep.* 20, 3057–3070.
61. Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., Vrieling, J.A.F.O., Bos, A.J., Drost, J., Rooijers, K., et al. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 149, 538–553.
62. Gennarino, V.A., Alcott, C.E., Chen, C.-A., Chaudhury, A., Gillentine, M.A., Rosenfeld, J.A., Parikh, S., Wheless, J.W., Roeder, E.R., Horovitz, D.D.G., et al. (2015). NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *Elife* 4. 10.7554/eLife.10782.
63. Grassi, E., Santoro, R., Umbach, A., Grosso, A., Oliviero, S., Neri, F., Conti, L., Ala, U., Provero, P., DiCunto, F., et al. (2018). Choice of alternative polyadenylation sites, mediated by the RNA-binding protein Elavl3, plays a role in differentiation of inhibitory neuronal progenitors. *Front. Cell. Neurosci.* 12, 518.
64. Ogorodnikov, A., Levin, M., Tattikota, S., Tokalov, S., Hoque, M., Scherzinger, D., Marini, F., Poetsch, A., Binder, H., Macher-Göppinger, S., et al. (2018). Transcriptome 3'end

- organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma. *Nat. Commun.* **9**, 5331.
65. Ha, K.C.H., Blencowe, B.J., and Morris, Q. (2018). QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data. *Genome Biol.* **19**. 10.1186/s13059-018-1414-4.
  66. Fang, Z., and Li, S. (2021). Alternative polyadenylation-associated loci interpret human traits and diseases. *Trends Genet.* **37**, 773–775.
  67. Mittleman, B.E., Pott, S., Warland, S., Zeng, T., Mu, Z., Kaur, M., Gilad, Y., and Li, Y. (2020). Alternative polyadenylation mediates genetic regulation of gene expression. *Elife* **9**. 10.7554/eLife.57492.
  68. Cui, Y., Arnold, F.J., Peng, F., Wang, D., Li, J.S., Michels, S., Wagner, E.J., La Spada, A.R., and Li, W. (2023). Alternative polyadenylation transcriptome-wide association study identifies APA-linked susceptibility genes in brain disorders. *Nat. Commun.* **14**, 583.
  69. Li, L., Huang, K.-L., Gao, Y., Cui, Y., Wang, G., Elrod, N.D., Li, Y., Chen, Y.E., Ji, P., Peng, F., et al. (2021). An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.* **53**, 994–1005.
  70. Liang, D., Elwell, A.L., Aygün, N., Krupa, O., Wolter, J.M., Kyere, F.A., Lafferty, M.J., Cheek, K.E., Courtney, K.P., Yusupova, M., et al. (2021). Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Nat. Neurosci.* **24**, 941–953.
  71. Flati, T., Gioiosa, S., Chillemi, G., Mele, A., Oliverio, A., Mannironi, C., Rinaldi, A., and Castrignanò, T. (2020). A gene expression atlas for different kinds of stress in the mouse brain. *Sci Data* **7**, 437.
  72. Matoba, N., Le, B.D., Valone, J.M., Wolter, J.M., Mory, J., Liang, D., Aygün, N., Broadaway, K.A., Bond, M.L., Mohlke, K.L., et al. (2023). WNT activity reveals context-specific genetic effects on gene regulation in neural progenitors. *bioRxiv*. 10.1101/2023.02.07.527357.
  73. Wulff, B.-E., and Nishikura, K. (2010). Substitutional A-to-I RNA editing. *Wiley Interdiscip. Rev. RNA* **1**, 90–101.
  74. Picardi, E., D’Erchia, A.M., Lo Giudice, C., and Pesole, G. (2017). REDportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **45**, D750–D757.
  75. Picardi, E., Manzari, C., Mastropasqua, F., Aiello, I., D’Erchia, A.M., and Pesole, G. (2015). Profiling RNA editing in human tissues: towards the inosinome Atlas. *Sci. Rep.* **5**, 14941.
  76. Roth, S.H., Levanon, E.Y., and Eisenberg, E. (2019). Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat. Methods* **16**, 1131–1138.
  77. Oakes, E., Anderson, A., Cohen-Gadol, A., and Hundley, H.A. (2017). Adenosine Deaminase That Acts on RNA 3 (ADAR3) Binding to Glutamate Receptor Subunit B Pre-mRNA Inhibits RNA Editing in Glioblastoma. *J. Biol. Chem.* **292**, 4326–4335.

78. Raghava Kurup, R., Oakes, E.K., Manning, A.C., Mukherjee, P., Vadlamani, P., and Hundley, H.A. (2022). RNA binding by ADAR3 inhibits adenosine-to-inosine editing and promotes expression of immune response protein MAVS. *J. Biol. Chem.* *298*, 102267.
79. Patil, V., Pal, J., Mahalingam, K., and Somasundaram, K. (2020). Global RNA editome landscape discovers reduced RNA editing in glioma: loss of editing of gamma-amino butyric acid receptor alpha subunit 3 (GABRA3) favors glioma migration and invasion. *PeerJ* *8*, e9755.
80. Khoshbakht, S., Beheshtian, M., Fattahi, Z., Bazazzadegan, N., Parsimehr, E., Fadaee, M., Vazehan, R., Faraji Zonooz, M., Abolhassani, A., Makvand, M., et al. (2021). and ; Genes with Ciliary Functions Cause Intellectual Disability in Multiple Families. *Arch. Iran. Med.* *24*, 364–373.
81. Badv, R.S., Mahdiannasser, M., Rasoulinezhad, M., Habibi, L., and Rashidi-Nezhad, A. (2022). CEP104 gene may involve in the pathogenesis of a new developmental disorder other than joubert syndrome. *Mol. Biol. Rep.* *49*, 7231–7237.
82. Srour, M., Hamdan, F.F., McKnight, D., Davis, E., Mandel, H., Schwartzenruber, J., Martin, B., Patry, L., Nassif, C., Dionne-Laporte, A., et al. (2015). Joubert Syndrome in French Canadians and Identification of Mutations in CEP104. *Am. J. Hum. Genet.* *97*, 744–753.
83. Hoffman, Y., Bublik, D.R., Ugalde, A.P., Elkon, R., Biniashvili, T., Agami, R., Oren, M., and Pilpel, Y. (2016). 3'UTR Shortening Potentiates MicroRNA-Based Repression of Pro-differentiation Genes in Proliferating Human Cells. *PLoS Genet.* *12*, e1005879.
84. Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* *17*, 761–772.
85. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 7028–7033.
86. Kobayashi, H., Saragai, S., Naito, A., Ichio, K., Kawauchi, D., and Murakami, F. (2015). Calm1 signaling pathway is essential for the migration of mouse precerebellar neurons. *Development* *142*, 375–384.
87. Bae, B., Gruner, H.N., Lynch, M., Feng, T., So, K., Oliver, D., Mastick, G.S., Yan, W., Pieraut, S., and Miura, P. (2020). Elimination of Calm1 long 3'-UTR mRNA isoform by CRISPR-Cas9 gene editing impairs dorsal root ganglion development and hippocampal neuron activation in mice. *RNA* *26*, 1414–1430.
88. Yang, Y.-C.T., Di, C., Hu, B., Zhou, M., Liu, Y., Song, N., Li, Y., Umetsu, J., and Lu, Z.J. (2015). CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics* *16*, 51.
89. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 9440–9445.
90. Aygün, N., Liang, D., Crouse, W.L., Keele, G.R., Love, M.I., and Stein, J.L. (2023). Inferring cell-type-specific causal gene regulatory networks during human neurogenesis. *Genome Biol.* *24*, 130.

91. Crouse, W.L., Keele, G.R., Gastonguay, M.S., and Churchill, G.A. (2021). A Bayesian model selection approach to mediation analysis. *bioRxiv*.
92. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443.
93. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Linnér, R.K., et al. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nat. Genet.* **50**, 1112.
94. Hofacker, I.L. (2009). RNA secondary structure analysis using the Vienna RNA package. *Curr. Protoc. Bioinformatics Chapter 12*, 12.2.1–12.2.16.
95. Enomoto, A., Murakami, H., Asai, N., Morone, N., Watanabe, T., Kawai, K., Murakumo, Y., Usukura, J., Kaibuchi, K., and Takahashi, M. (2005). Akt/PKB regulates actin organization and cell motility via Girdin/APE. *Dev. Cell* **9**, 389–402.
96. Enomoto, A., Asai, N., Namba, T., Wang, Y., Kato, T., Tanaka, M., Tatsumi, H., Taya, S., Tsuboi, D., Kuroda, K., et al. (2009). Roles of disrupted-in-schizophrenia 1-interacting protein girdin in postnatal development of the dentate gyrus. *Neuron* **63**, 774–787.
97. Shimizu, Y., and Kawasaki, T. (2021). Histone acetyltransferase EP300 regulates the proliferation and differentiation of neural stem cells during adult neurogenesis and regenerative neurogenesis in the zebrafish optic tectum. *Neurosci. Lett.* **756**, 135978.
98. Lundin, E., Wu, C., Widmark, A., Behm, M., Hjerling-Leffler, J., Daniel, C., Öhman, M., and Nilsson, M. (2020). Spatiotemporal mapping of RNA editing in the developing mouse brain using in situ sequencing reveals regional and cell-type-specific regulation. *BMC Biol.* **18**, 6.
99. Jiang, K., Koob, J., Chen, X.D., Krajeski, R.N., Zhang, Y., Volf, V., Zhou, W., Sgrizzi, S.R., Villiger, L., Gootenberg, J.S., et al. (2022). Programmable eukaryotic protein synthesis with RNA sensors by harnessing ADAR. *Nat. Biotechnol.* [10.1038/s41587-022-01534-5](https://doi.org/10.1038/s41587-022-01534-5).
100. Qian, Y., Li, J., Zhao, S., Matthews, E.A., Adoff, M., Zhong, W., An, X., Yeo, M., Park, C., Yang, X., et al. (2022). Programmable RNA sensing for cell monitoring and manipulation. *Nature* **610**, 713–721.
101. Hassan, M.Q., Gordon, J.A.R., Lian, J.B., van Wijnen, A.J., Stein, J.L., and Stein, G.S. (2010). Ribonucleoprotein immunoprecipitation (RNP-IP): a direct in vivo analysis of microRNA-targets. *J. Cell. Biochem.* **110**, 817–822.
102. Ma, Y., Dammer, E.B., Felsky, D., Duong, D.M., Klein, H.-U., White, C.C., Zhou, M., Logsdon, B.A., McCabe, C., Xu, J., et al. (2021). Atlas of RNA editing events affecting protein expression in aged and Alzheimer’s disease human brain tissue. *Nat. Commun.* **12**, 7035.
103. Robins, C., Liu, Y., Fan, W., Duong, D.M., Meigs, J., Harerimana, N.V., Gerasimov, E.S., Dammer, E.B., Cutler, D.J., Beach, T.G., et al. (2021). Genetic control of the human brain proteome. *Am. J. Hum. Genet.* **108**, 400–410.

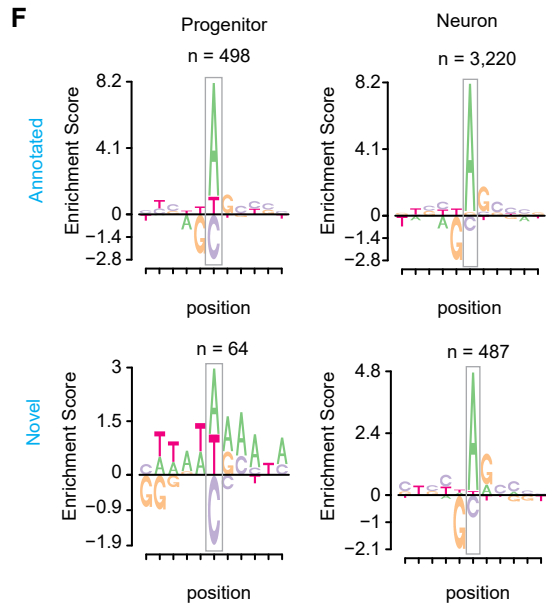
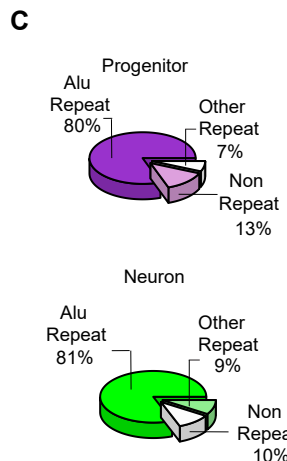
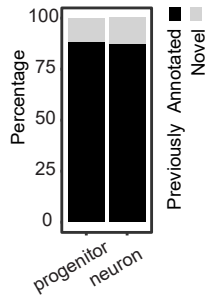
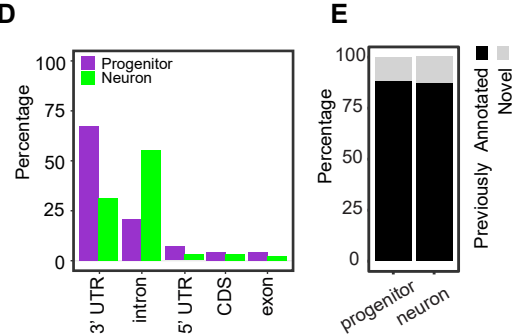
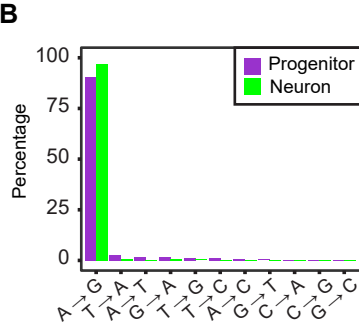
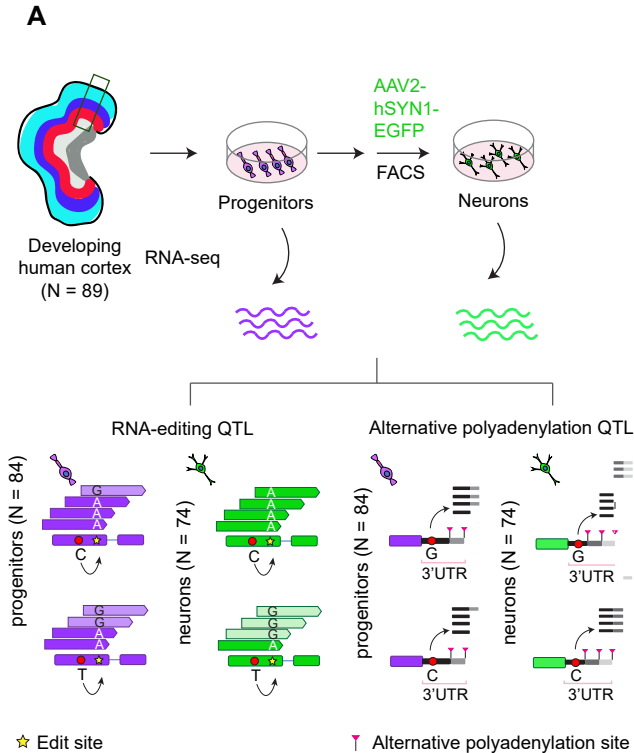
104. Robinson, J.W., Battram, T., Baird, D.A., Haycock, P.C., Zheng, J., Hemani, G., Chen, C.-Y., and Gaunt, T.R. (2022). Evaluating the potential benefits and pitfalls of combining protein and expression quantitative trait loci in evidencing drug targets. *bioRxiv*. 10.1101/2022.03.15.484248.
105. He, B., Shi, J., Wang, X., Jiang, H., and Zhu, H.-J. (2020). Genome-wide pQTL analysis of protein expression regulatory networks in the human liver. *BMC Biol.* 18, 97.
106. Stein, J.L., de la Torre-Ubieta, L., Tian, Y., Parikshak, N.N., Hernández, I.A., Marchetto, M.C., Baker, D.K., Lu, D., Hinman, C.R., Lowe, J.K., et al. (2014). A quantitative framework to evaluate modeling of cortical development by neural stem cells. *Neuron* 83, 69–86.
107. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12.
108. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
109. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* 91, 839–848.
110. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
111. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299.
112. van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* 12, 1061–1063.
113. Dey, K.K., Xie, D., and Stephens, M. (2018). A new sequence logo plot to highlight enrichment and depletion. *BMC Bioinformatics* 19, 473.
114. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419.
115. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
116. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300.
117. Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* 18, 100–106.

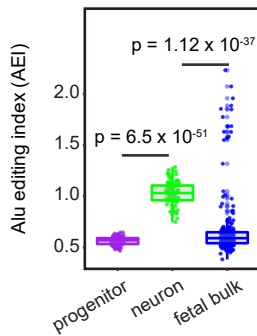
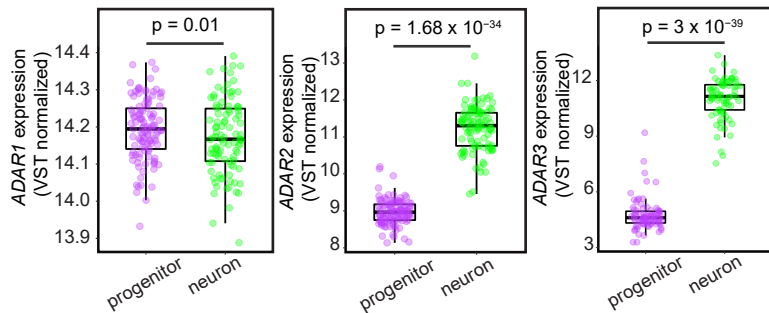
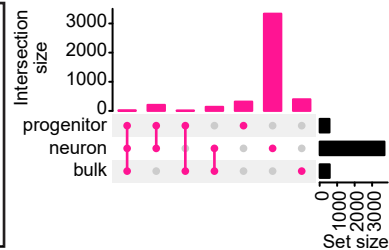
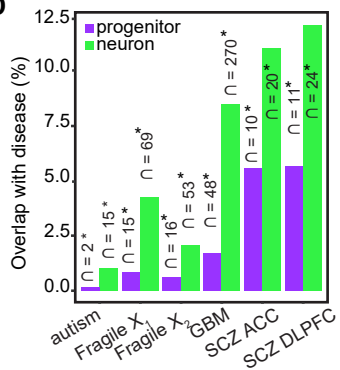
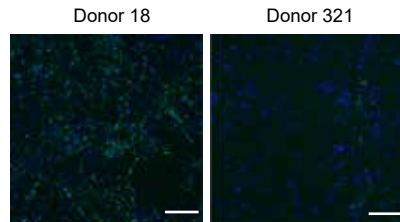
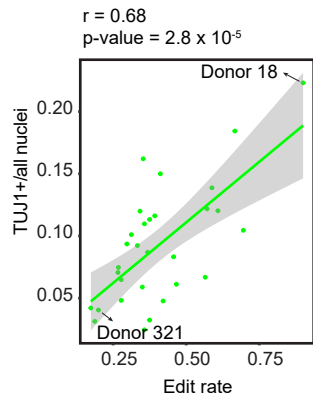
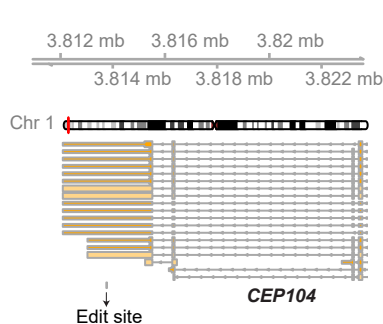


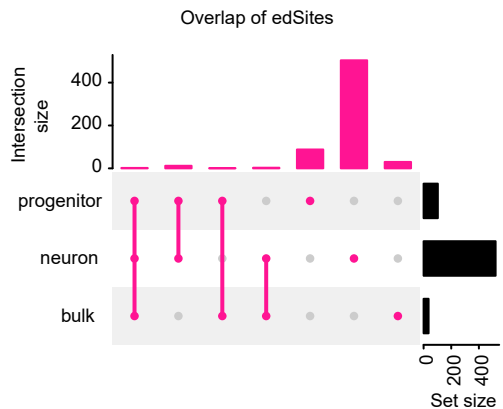
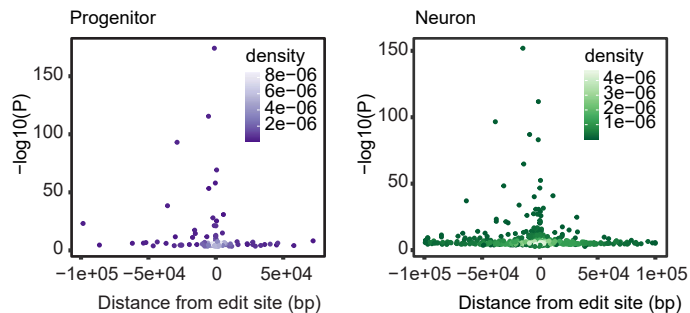
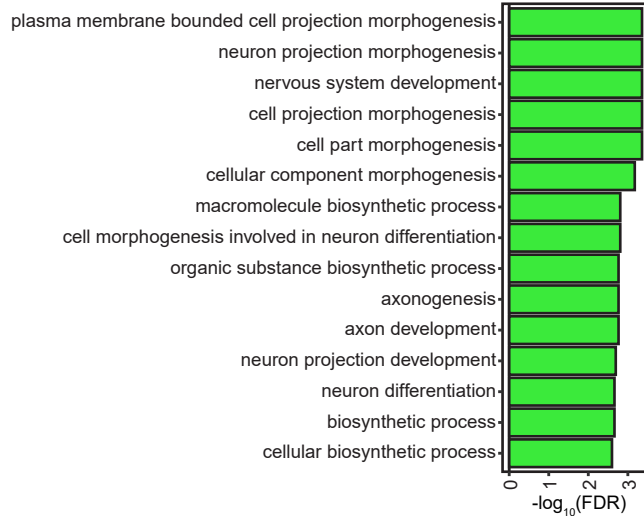
118. Huang, Q.Q., Ritchie, S.C., Brozynska, M., and Inouye, M. (2018). Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic Acids Research* 46, e133–e133. [10.1093/nar/gky780](https://doi.org/10.1093/nar/gky780).
119. Davis, J.R., Fresard, L., Knowles, D.A., Pala, M., Bustamante, C.D., Battle, A., and Montgomery, S.B. (2016). An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am. J. Hum. Genet.* 98, 216–224.
120. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348–354.
121. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46, 100–106.
122. Dabney, A., Storey, J.D., and Warnes, G.R. (2010). qvalue: Q-value estimation for false discovery rate control. R package version 1.
123. Do, C.B., Woods, D.A., and Batzoglou, S. (2006). CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* 22, e90–e98.
124. Danaee, P., Rouches, M., Wiley, M., Deng, D., Huang, L., and Hendrix, D. (2018). bpRNA: large-scale automated annotation and analysis of RNA secondary structure. *Nucleic Acids Res.* 46, 5381–5394.
125. Civelek, M., Wu, Y., Pan, C., Raulerson, C.K., Ko, A., He, A., Tilford, C., Saleem, N.K., Stančáková, A., Scott, L.J., et al. (2017). Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *Am. J. Hum. Genet.* 100, 428–443.
126. Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* 50, 668–681.
127. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D., and van der Sluis, S. (2018). Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat. Commun.* 9, 905.
128. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* 51, 63–75.
129. Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., et al. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* 18, 1091–1102.
130. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* 51, 404–413.







**A****B****C****D****E**

**A****B****C****D**