

Title: Expansion of RNA sequence diversity and RNA editing rates throughout human cortical development

AUTHORS: Ryn Cudleston^{1,2,3,4}, Laura Sloofman^{1,2,3,4‡}, Lindsay Liang^{5‡}, Enrico Mossotto^{1‡}, Xuanjia Fan^{1,2,3}, Minghui Wang^{2,6,7}, Bin Zhang^{2,6,7}, Jiebiao Wang⁸, Nenad Sestan^{9,10,11,12,13}, Bernie Devlin¹⁴, Kathryn Roeder¹⁵, Joseph D. Buxbaum^{1,2,3,4,16}, Stephan J. Sanders⁵, Michael S. Breen^{1,2,3,4,16#}

AFFILIATIONS:

¹Department of Psychiatry at Mount Sinai, New York, New York, 10029, USA

²Department of Genetics and Genomic Sciences at Mount Sinai, New York, New York, 10029, USA

³Seaver Autism Center for Research and Treatment at Mount Sinai, New York, New York, 10029, USA

⁴Pamela Sklar Division of Psychiatric Genomics at Mount Sinai, New York, New York, 10029, USA

⁵Department of Psychiatry and Behavioral Sciences and UCSF Weill Institute for Neurosciences, University of California, San Francisco, CA 94158, USA

⁶Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York 10029, NY, USA.

⁷Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

⁸Department of Biostatistics, University of Pittsburgh, 130 De Soto Street, Pittsburgh, Pennsylvania 15261, USA

⁹Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT 06510, USA

¹⁰Program in Cellular Neuroscience, Neurodegeneration, and Repair and Yale Child Study Center, Yale School of Medicine, New Haven, CT 06510, USA

¹¹Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06520, USA

¹²Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA

¹³Department of Comparative Medicine, Program in Integrative Cell Signaling and Neurobiology of Metabolism, Yale School of Medicine, New Haven, CT 06510, USA

¹⁴Department of Psychiatry, University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15213, USA

¹⁵Carnegie Mellon University, Statistics & Data Science Department, Pittsburgh, Pennsylvania 15213, USA

¹⁶Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, 10029, USA

[‡]Individuals share equal contribution.

[#]Corresponding author. Email: michael.breen@mssm.edu

ABSTRACT

Post-transcriptional modifications by RNA editing are essential for neurodevelopment, yet their developmental and regulatory features remain poorly resolved. We constructed a full temporal view of base-specific RNA editing in the developing human cortex, from early progenitors through fully mature cells found in the adult brain. Developmental regulation of RNA editing is characterized by an increase in editing rates for more than 10,000 selective editing sites, shifting between mid-fetal development and infancy, and a massive expansion of RNA hyper-editing sites that amass in the cortex through postnatal development into advanced age. These sites occur disproportionately in 3'UTRs of essential neurodevelopmental genes. These profiles are preserved in non-human primate and murine models, illustrating evolutionary conserved regulation of RNA editing in mammalian cortical development. RNA editing levels are commonly genetically regulated (editing quantitative trait loci, edQTLs) consistently across development or predominantly during prenatal or postnatal periods. Both consistent and temporal-predominant edQTLs co-localize with risk loci associated with neurological traits and disorders, including attention deficit hyperactivity disorder, schizophrenia, and sleep disorders. These findings expand the repertoire of highly regulated RNA editing sites in the brain and provide insights of how epitranscriptional sequence diversity by RNA editing contributes to neurodevelopment.

MAIN

Over the course of development, DNA-encoded genes are precisely transcribed into functional RNA sequences, which undergo a variety co- and post-transcriptional modifications. Post-transcriptional modifications by RNA editing are a major contributor to the global diversity of RNA sequences^{1,2} and are predicted to occur at over a hundred million sites in the human transcriptome³⁻⁵, with most impacting neuronal genes^{6,7}. While RNA editing is essential for neurodevelopment, there is little understanding of the temporal dynamics of RNA editing during human brain development. For instance, it remains unclear to what degree the frequencies and rates of RNA editing sites are developmentally regulated and when and where these sites are established from early fetal through late postnatal periods. Moreover, the genetic regulation of RNA editing through brain development periods also remains poorly resolved.

Adenosine-to-Inosine (A-to-I) editing is the most prevalent form of RNA editing in the brain, whereby inosine is recognized by the cellular machinery as guanosine during translation⁸. A-to-I editing can occur either at a single isolated adenosine, termed “selective editing”, or across many neighboring adenosines within an extended region or cluster, termed “RNA hyper-editing”⁹⁻¹¹, and both are driven by adenosine deaminases acting on RNA (ADAR) enzymes. The vast majority of RNA editing occurs in primate-specific Alu elements^{1,10}. These base-specific changes recode amino acid sequences of proteins, alter the locations of start or stop codons¹², influence neuronal splicing patterns¹³, affect the ability of miRNAs to bind to their target sites and can alter the stability of the RNA secondary structures^{14,15}. A requisite for neuronal development and function, several sites are known to drive essential physiological processes, including: tight regulation of Ca^{2+} permeability (*GRIA2*, Q/R)^{16,17}, acceleration of recovery rates from desensitization (*GRIA2-4*, R/G)¹⁸, actin cytoskeletal remodeling at excitatory synapses (*CYFIP2*, K/E)^{19,20}, gating kinetics of inhibitory receptors (*GABBAR*, I/M)^{21,22}, maintenance of action potential for voltage-dependent potassium channels ($\text{K}_v1.1$, I/V), and calmodulin binding mediating inhibitory Ca^{2+} -feedback on channels ($\text{Ca}_v1.3$).

Recognizing their importance, the developmental properties of recoding sites have been extensively examined in murine models of neurodevelopment and report a precise increase in editing levels from embryonic to postnatal periods²³⁻²⁷. It has been proposed that such editing activity is not solely explained by ADAR expression, but by the presence of a strict regulatory network that controls editing during development. Fewer studies have profiled editing activity across human brain development, and leverage either a small number of samples, a narrow developmental window, or study a select number of editing sites²⁸⁻³². Markedly, one report uncovered 742 RNA editing sites that increased in editing levels during postnatal periods³¹ and a separate investigation also found an increasing pattern of global editing activity across *in vitro* maturation of hiPSC-derived brain organoids³²; offering a partial view of RNA editing throughout neurodevelopment. Still, the developmental and regulatory features of RNA editing in human brain have not been fully compiled. Such information is essential to understand the biological roles of RNA editing throughout the widely different neurodevelopmental stages, from early differentiating cellular populations to the broad range of structures found in the adult brain.

To address these gaps, we present a systematic evaluation of both selective RNA editing and RNA hyper-editing in the developing human cortex. Our study covers a wide-range of developmentally distinct samples (**Figure 1**). We anchor our analysis around BrainVar data³³, which comprises paired whole-genome and bulk tissue RNA-sequencing (RNA-seq) data from the dorsolateral prefrontal cortex (DLPFC) of 176 donors spanning prenatal and postnatal developmental periods, from six post-conception weeks (PCWs) to young adulthood (20 years). The

frontal cerebral wall was assayed in nine brains before ten PCWs. These data were integrated with *in vitro* human embryonic and induced pluripotent stem cell models of corticogenesis and hundreds of cortical samples from advanced ages. We identify a dramatic shift in selective RNA editing and hyper-editing between mid-fetal development and infancy, which refines the timing of these events and the degree to which each editing site is involved. Our analytic framework illustrates evolutionary conservation of these profiles in the non-human primate and murine models of neurodevelopment and elucidates interactions between RNA editing sites and genes essential for neuronal differentiation, mRNA metabolism, synaptic signaling, receptor sequences and those important for neurodevelopment. We also uncover the genetic regulation of RNA editing and identify hundreds of *cis*-editing quantitative trait loci (edQTLs) and classify their effects as prenatal-predominant, postnatal-predominant, or constant across brain development. Several edQTLs co-localize with risk variants for neuropsychiatric and neurodevelopmental disorders. Data and analyses we present here refine our understanding of the RNA editing-mediated mechanism throughout cortical development and will provide significant benefits to the fields of neuroscience, genomics, and medicine.

Increased rate of selective editing through human cortical development

We computed an Alu editing index (AEI) as a measure of global selective A-to-G editing activity for each individual defined as the total number of selective A-to-G mismatches over the total coverage of adenosines in Alu elements (**Supplemental Table 1**, **Figure S1**). A threefold increase in global selective A-to-G editing rates occurred during cortical development ($p=2.2\times10^{-28}$, Cohen's $d=2.55$), with a major shift between the mid-fetal period and infancy (**Figure 2A**). In parallel, *ADAR1* ($p=2.9\times10^{-12}$), *ADAR2* ($p=3.1\times10^{-5}$), and *ADAR3* ($p=0.009$) showed strong postnatal bias in expression (**Figure 2B**). Cellular deconvolution of bulk RNA-seq profiles revealed a decrease in the proportions of immature/fetal neurons and an increase in mature neurons from prenatal to postnatal periods (**Figure 2C**). Variation in the AEI was largely explained by age ($R^2=4.2\%$), followed by proportions of mature neurons within a sample ($R^2=3.3\%$), *ADAR2* ($R^2=0.49\%$) and *ADAR1* expression ($R^2=0.44\%$) (**Figure S2**). To provide deeper cell-specific information, we turned to single-nuclei RNA-seq of the DLPFC covering mid-fetal and postnatal periods. These data confirm ADAR expression levels are postnatally biased and that both *ADAR1* and *ADAR2* are highly expressed in mature excitatory neurons, whereas *ADAR3* shows a strong preference for a small subclass of inhibitory neurons (**Figure S3**). In keeping with the estimates of AEI variation, a significant postnatal bias in global editing activity remained after correcting for the proportion of mature neurons ($p=2.1\times10^{-6}$).

We juxtaposed global selective A-to-G editing rates from BrainVar prenatal periods (Epoch 1) with two *in vitro* models of corticogenesis: 1) a human embryonic stem cell (hESC) model, spanning early stages of pluripotency (day 0) through upper layer generation (day 77); and 2) a human induced pluripotent stem cell (hiPSC) model, including 127 samples covering early differentiating cells (days 2-9) through mature differentiated neurons (day 77) (**Supplemental Table 1**). Both hESCs ($p=3.9\times10^{-5}$, Cohen's $d=1.44$) and hiPSCs ($p=3.5\times10^{-24}$, Cohen's $d=1.67$) captured this increase in global selective A-to-G editing (**Figure 2D**), and these rates were comparable across the two models. Notably, editing rates during late differentiation stages (days 49-77) in hiPSCs were similar to those in fetal brain tissue (Epoch 1) of BrainVar ($p=0.38$) (**Figure S4**).

Next, global selective A-to-G editing rates from BrainVar postnatal periods (Epoch 3) were contrasted with 261 cortical samples from older adults (61-108 years) (**Figure S5, Supplemental Table 1**). While the AEI was not dynamically regulated throughout advanced aging ($p=0.39$) (**Figure 2E**), editing activity steadily increased

compared to Epoch 3 of BrainVar ($p=3.1\times10^{-5}$) (**Figure 2F**), suggesting that cortical global editing rates in the cortex might peak between 20-59 years of life. To test this, we explored the AEI in a smaller independent dataset of bulk forebrain ($n=55$) and cerebellar ($n=59$) tissues spanning fetal and postnatal periods (4 PCWs to 58 years) (**Supplemental Table 1**). Both the bulk forebrain and cerebellar tissues confirmed the increased rate of editing throughout development ($p=6.2\times10^{-16}$, Cohen's $d=2.97$; $p=1.2\times10^{-12}$, Cohen's $d=2.44$, respectively) and peak during young adulthood (~30 years) (**Figure 2G**). We also show that editing activity in the brain displays the strongest developmental increase relative to four other human tissues sampled across the lifespan (**Figure S6**). Overall, global editing activity rises throughout maturation from early progenitors through fully differentiated neurons in the adult brain, even after adjustment for neuronal proportions (**Figure 2G**).

Systematic characterization of selective editing events

To catalogue high-confidence selective RNA editing sites, a series of detection-based thresholds were imposed and all common and private genomic variants were removed (*see Methods*) (**Figure S7**). Using the resulting high-quality RNA editing sites and a sliding window spanning three neighboring developmental periods, a uniform number of sites per developmental window was observed (**Figure 3A**). RNA editing sites within each window were then compared to all remaining windows and a convergence of sites consistently edited either during prenatal or postnatal developmental periods was identified (**Figure 3B**). Based on this result, we aggregated RNA editing sites across i) prenatal samples only, ii) postnatal samples only, and iii) all available samples. We identified 3,879 prenatal predominant sites, 6,969 postnatal predominant sites and 10,652 'common' sites with high detection rates across all cortical samples (**Figure 3C, Supplemental Table 2**). The majority of postnatal predominant and common sites were edited with ~25% editing efficiently, while editing rates of prenatal predominant sites were significantly higher (**Figure 3D**).

Validating the accuracy of our approach, these sites showed consistent hallmark signatures of RNA editing: i) The vast majority were A-to-G editing events (69% prenatal, 89% postnatal), mapped to Alu repeats (71.2%) and were known sites catalogued in editing databases (70.3%) (**Figure 3E**); ii) We confirmed the local sequence motif for all A-to-G editing events in which guanosine was depleted -1bp upstream and enriched +1 downstream the targeted adenosine (**Figure 3F**); iii) The majority of sites occurred in 3'UTRs and introns with few sites in coding regions (**Figure 3G-I**), with a significant enrichment of 3'UTR editing during postnatal periods ($p=2.7\times10^{-6}$) (**Figure S8**); iv) Most prenatal predominant sites mapped to genes with a falling expression trajectory over development (77.4%), while postnatal predominant sites mapped to genes with a rising expression (54.4%) (**Figure 3J**); v) We confirmed the temporal specificity for ~45% prenatal predominant sites ($n^{sites}=1,745$) in hESC and hiPSC models of corticogenesis (which were absent during advanced age) and ~89% of postnatal predominant sites ($n^{sites}=6,132$) in advanced age (which were absent hESCs and hiPSCs) (**Figure 3K-L**).

Developmental regulation of selective editing events

For the 10,562 *bona fide* selective RNA editing sites in common across all BrainVar samples, differences associated with developmental age had the largest genome-wide effect on editing rates, followed by ADAR expression and neuronal proportions (**Figure S9**). Principal components accurately distinguished prenatal from postnatal samples based on editing rates for these sites (**Figure 4A**). Differential editing identified developmentally-regulated sites: 6,450 showed higher editing postnatally ('postnatal biased'), 776 prenatally ('prenatal biased'), and 3,426 were constant ('unbiased') across cortical development (**Figure 4B, Supplemental**

Table 3). While these results were robust to adjusting for mature neuronal cell proportions, adjusting for *ADAR1* and *ADAR2* resulted in twofold reduction of developmentally regulated sites (**Figure S9, Supplemental Table 3**). Notably, the majority of postnatally biased RNA editing sites (~86%) were known, while the majority of prenatally biased RNA editing sites (~81%) were novel, despite similar sample numbers and the same thresholds for detection (**Figure 4C**). Novel sites occurred across all genic regions in a distribution similar to known sites and showed significantly more supporting read coverage compared to known sites (**Figure S10**).

Most developmentally regulated editing sites increased in editing efficiencies and mapped to 3'UTRs (**Figure 4D**). Accordingly, we modeled whether RNA editing in 3'UTRs might influence miRNA binding efficiency (*see Materials and Methods*). A significant reduction in miRNA binding energy was observed in edited relative to un-edited 3'UTRs ($p=2.3\times 10^{-32}$, Mann-Whitney U test) (**Figure S11**), suggesting a regulatory role for RNA editing in cortical development. Moreover, concordance analysis between RNA editing and corresponding gene expression profiles identified editing in 3'UTRs as having the largest overall effect and negatively correlated with gene expression across development (**Figure S12**).

To infer the biology of developmentally regulated events, unsupervised weighted correlation analysis identified four discrete groups of co-regulated editing modules (M) that varied in the number of editing sites, gene content and editing efficiencies (**Figure 4E, Figure S13**). All modules were enriched for general terms related to neuronal projection, synapses, axons and dendrites, but were also enriched for unique functional processes (**Figure 4F, Supplemental Table 3**). The majority of sites mapped to M1 ($n^{sites}=10,038$), which was postnatally biased ($R=0.87$, $p=4.0\times 10^{-56}$), but implicated in biological processes known to be prenatally biased in gene expression, including neurogenesis, mRNA metabolism, cytoskeletal protein binding and 3'UTR binding. M2 ($n^{sites}=204$), the grey module, was prenatally biased ($R=-0.78$, $p=2.0\times 10^{-36}$) and associated with neuronal projection, cell signaling, cell adhesion molecule binding and neurotransmitter binding and receptor activity. M3 ($n^{sites}=116$), the smallest module was also prenatally biased ($R=-0.33$, $p=1.0\times 10^{-5}$) and implicated in synaptic transmission, fatty acid biosynthesis, oxidoreductase activity and cadherin binding. M4 ($n^{sites}=258$) was the module with the lowest level of editing activity, also prenatally biased ($R=-0.73$, $p=2.0\times 10^{-30}$), and was implicated in processes related to ion transporter activity to synapse, regulation of adherens junction organization and calcium-dependent protein binding. Notably, M1 was significantly enriched for: *i*) genes that confer risk for autism spectrum disorder (ASD), schizophrenia and educational attainment; *ii*) differentially expressed genes in cortical samples from individuals with ASD; and *iii*) dysregulated RNA editing sites in cortical samples from individuals with schizophrenia and sites related to postnatal development (**Figure 4G-H**). Furthermore, supervised analysis of candidate gene sets confirmed that RNA editing had an overall negative association with neurogenesis, neuronal differentiation, mRNA metabolism (genes with high prenatal expression) and a positive association with synaptic signaling (genes with high postnatal expression) (**Figure S14**).

We validated the temporal trajectories for the vast majority of developmentally regulated sites in two smaller independent developmental brain transcriptome datasets, and observed high transcriptome-wide concordance of delta editing rates across forebrain ($R^2=0.58$) and cerebellum development ($R^2=0.31$) (**Figure S15**). Next, in 261 aged cortical samples, ~90% of all common editing sites with sufficient coverage ($n^{sites}=9,568$) were reproduced and these data re-confirmed that M1 was consistently highly edited during advanced aging, whereas the remaining modules were edited to a lesser extent (~0-2% editing efficiency) (**Figure S16A-C, Supplemental Table 4**). In addition, ~20% of all common editing events ($n^{sites}=2,168$) were re-confirmed in hESCs and validated prenatally

biased modules M2 and M3, which contained significantly more coverage and higher editing rates relative to modules M1 and M4 (**Figure S16D-F, Supplemental Table 4**).

Dynamic developmental regulation of RNA recoding sites

RNA editing can recode amino acids and a number of recoding events play an important role for neuronal function and development. To expand upon this current knowledge, we observed that recoding sites displayed the smallest developmental changes in editing efficiencies compared to sites in other genic regions, as expected (**Figure 5A**). Second, recoding events were more highly conserved relative to sites in all other genic regions (**Figure 5B**) and enriched for components related to glutamate receptor complex (FDR $p=4.3\times10^{-6}$), ion-channel activity (FDR $p=1.9\times10^{-3}$), cadherin binding (FDR $p=0.002$), and mRNA metabolism (FDR $p=0.002$) (**Figure S17, Table S3**). Third, we identified 32 recoding events that displayed significant developmental changes in editing efficiencies (**Figure 5C-E**), including nine high-confidence novel recoding events that validate in advanced age (**Figure S18**). Ranking these sites by developmental effect sizes re-confirmed several recoding sites that are known to increase in editing efficiency through cortical development, including those mapping to a collection of excitatory, inhibitory and G-coupled protein receptors (e.g., *GRIK2*, *GABRA3*, *GRIA2*, *CYFIP2*). However, we also highlight numerous known recoding sites with previously undefined roles in cortical development: an arginine/glycine conversion (R/G) on the 3rd exon of cyclin I (*CCNI*); an I/M and S/G site on the 3rd exon of signal recognition particle 9 (*SRP9*); an R/G site on the 3rd exon of SON DNA and RNA binding protein (*SON*, the cause of ZTTK syndrome); a K/R site on the 6th exon of endonuclease VIII-like 1 (*NEIL*); a T/A site on the 6th exon of EEF1A lysine methyltransferase 2 (*EEF1AKMT2*), among others (**Figure 5D-E**). Finally, of the nine novel recoding events, seven were prenatally biased in their editing efficiencies and two were C-to-T edits, including a Q/R site on the 2nd exon of ribosomal protein S6 kinase C1 (*RPS6KC1*), D/G site on the 11th exon of MINDY lysine 48 deubiquitinase 3 (*MINDY3*), and an R/G site on the 12th exon of sorting nexin 2 (*SNX2*), among others.

RNA hyper-editing through human cortical development

Using BrainVar data, we first measured RNA hyper-editing across all substitution types and A-to-G accounted for ~99% of edits (**Figure S19**). Therefore, this class was the primary focus of ensuing analyses. Hyper-editing in 3'UTRs and introns comprised ~80% of all events with very few events in coding regions (~0.24%). Hyper-editing occurred in clusters of ~80bp in length and each cluster contains ~10.8 editing sites (**Supplemental Table 5**). The rate of hyper-editing significantly increased from prenatal ($\mu=24,428$ sites) through postnatal ($\mu=78,441$ sites) periods ($p=1.4\times10^{-17}$) (**Figure 6A**). To minimize technical variability and enable a direct comparison of hyper-editing across development, we normalized the rate of hyper-editing to the number of million mapped bases per sample. Normalized hyper-editing rates increased in frequency into postnatal development ($p=7.2\times10^{-17}$, Cohen's $d=3.06$) (**Figure 6B**). These rates were associated with the proportion of mature neurons ($R^2=0.41$) and ADAR expression (**Figure S20**), and remained significantly postnatally biased following adjustment for neuronal cell type frequencies ($p=1.0\times10^{-11}$). The local sequence motif for all A-to-G hyper-editing sites was consistent with the expected distribution of guanosines upstream and downstream the target adenosine (**Figure 6C**).

We identified 1,754 genes that amassed a significant number of A-to-G hyper-editing sites throughout development. After adjusting for the proportion of mature neurons and gene length, we retained 737 genes enriched with postnatal hyper-editing signal (**Figure 6D-E, Supplemental Table 5**). Approximately 74% displayed developmental expression profiles that are either falling or non-transient over development, suggesting

that the hyper-editing enrichment is largely independent of the corresponding gene expression levels (**Figure 6F**). Genes enriched with postnatal hyper-editing signal were also enriched for genes implicated in genetic risk for autism spectrum disorder, educational attainment, and schizophrenia, as well as genes that are significantly differentially expressed in the cortex of individuals with ASD (**Figure 6G**). For example, gradual accumulation of hyper-editing events over cortical development was observed in the 1st and 2nd introns of potassium voltage-gated channel interacting protein 4 (*KCNIP4*) and in the 4th, 10th and 13th introns of neurexin 3 (*NRXN3*). (**Figure 6H-I**).

These findings were compared with hyper-editing signatures during *in vitro* corticogenesis and in cortical samples from advanced ages (**Supplementary Table 5**). In hESCs, a significant number of hyper-editing sites accumulated throughout neuronal maturation and differentiation ($p=0.007$) (**Figure 6J**). Notably, the frequency of hyper-editing was significantly elevated ($\mu=21,796$ sites) during pluripotency, in contrast to the remaining days post neural induction ($\mu=2,163$ sites, $p=1.9\times 10^{-9}$), and a marked ~58% of these sites mapped to coding regions (**Figure S21**). In cortical samples during advanced age, the normalized hyper-editing rate was not dynamically regulated (**Figure 6K**) but increased twofold compared to the Epoch 3 of BrainVar ($p=5.7\times 10^{-11}$). All hyper-editing sites were enriched for a common sequence motif (**Figure S21-22**). Moreover, genes displaying a significant postnatal enrichment of RNA hyper-editing sites in BrainVar were validated during advanced age with high concordance ($R^2=0.58$) (**Figure S23**). Taken together, we observe a substantial enrichment of A-to-G hyper editing events that accumulate in the cortex through development (**Figure 6L**).

Finally, given that the majority of hyper-editing sites accumulate in introns of neurodevelopmental genes, we tested whether these sites (*i*) are predicted to be splice altering, (*ii*) occur in retained introns and/or (*iii*) occur in ultra-conserved non-coding elements (UCNEs), which can act as transcriptional regulators. First, the number of intronic RNA hyper-editing events predicted to be splice altering increased throughout development ($p=2.9\times 10^{-19}$), but ultimately accounted for a small subset of the total number of hyper-editing sites (**Figure S24**). Second, while intron retention (IR) was more commonly detected in postnatal periods (~57,311 introns, ~10,848 genes) relative to prenatal periods (~54,638 introns, ~10,176 genes) ($p=0.0006$), the percent intron inclusion ratio was moderately elevated during prenatal periods ($p=0.01$) (**Figure S25A-D, Supplemental Table 6**). Importantly, ~80% of genes enriched with postnatal intronic hyper-editing events display heightened IR during postnatal periods, but greater IR was not correlated with the expression of their host genes (**Figure S25E-F**). Third, a significant over-representation of postnatal hyper-editing signal was also detected in genes harboring UCNEs ($p=6.3\times 10^{-5}$), which might underlie suboptimal regulation during ageing (**Figure S23C**).

Conservation of increased selective editing and hyper-editing in mammalian cortical development

We tested whether these developmental patterns of RNA editing constitute a conserved regulatory mechanism involved in cortical physiology in two animal models of cortical development: *i*) bulk tissue RNA sequencing of four cortical regions of 26 rhesus macaques, from 60 post-conception days to 11 years; *ii*) whole-cortex RNA-sequencing of 18 wild-type mice, from embryonic day 14.5 to 21 postnatal months (**Supplemental Table 7**). In macaque, we observed a significant postnatal bias in both the AEI ($p=3.5\times 10^{-17}$) and in the frequency of RNA hyper-editing sites ($p=7.1\times 10^{-10}$) (**Figure 7A-C**). Hyper-editing clusters ranged in size from 65-72bp in length, each contained ~7.8 hyper-editing sites, and all sites shared a local sequence motif similar to that observed in humans (**Figure 7D**). These findings also reproduced in mouse, including the increased rate of selective RNA editing ($p=1.9\times 10^{-12}$), increased frequency of RNA hyper-editing sites ($p=1.8\times 10^{-7}$), and conservation of common

local sequence motif for all A-to-G hyper-editing events (**Figure 7E-H**). Furthermore, several developmentally regulated recoding sites validated across species, with few sites showing human-specific developmental regulation (**Figure S26**). Finally, we observed that global editing rates and the frequency of hyper-editing sites are highest in human cortex, followed by macaque and subsequently mouse (**Figure 7I**). Importantly, all hyper-editing sites in the current study were robust to potential false positives and were not confounded by common genomic variation (**Figure S27**).

Developmental cis-edQTLs and co-localization with neurological traits and disorders

WGS data were used to detect SNPs that could have an effect on RNA editing levels (edQTL, editing quantitative trait loci). RNA editing levels were fit to SNP genotypes, covarying for developmental period, sex, the first five principal components of ancestry, as well as *ADAR1* and *ADAR2* expression. This permitted identification of editing levels under tight genetic control, irrespective of the corresponding ADAR activity. To distinguish temporal-predominant edQTLs, we performed three *cis*-edQTL analyses: i) prenatal samples only ($n=116$, periods 1-7); ii) postnatal samples only ($n=60$, periods 8-12); and iii) all samples ($n=176$, periods 1-12). We defined a 1Mb window (\pm) to search for SNP-editing pairs of an editing site and identified 10,803 *cis*-edQTLs from all samples, 5,386 *cis*-edQTLs from prenatal samples, and 1,008 *cis*-edQTLs from postnatal samples at FDR < 5% (**Supplemental Table 8**). These edQTLs comprised a total of 1039, 1,008 and 802 unique editing sites (eSites) across all samples, prenatal samples and postnatal samples, respectively, and each lead SNP was located close to their associated editing site (15kb \pm nt) (**Figure 8A-B**).

The majority of edQTLs were consistent across development, reaching nominal significance in all three analyses with the same direction of effect (~77%, **Figure 8C**). Many edQTLs were prenatal-predominant, with greater prenatal than postnatal effect sizes, while fewer edQTLs were postnatal-predominant, with significantly greater postnatal than prenatal effect sizes. Regardless of their temporal predominance, a substantial fraction of all corresponding eSites increased in editing efficiency through development (~95%), enriched in module M1, and mapped to 3'UTRs and introns (**Figure 8D-E**). edQTL SNPs (edSNPs) were also examined for association with gene expression levels by calculating the overlap between edQTLs and previously computed expression QTL (eQTL) summary statistics. Overall, a total of 969 edSNPs (12%) were also associated with variation in gene expression, for which 27% were associated with a gene and one or more editing sites within the same gene.

Finally, we queried consistent and temporal predominant edQTLs for co-localization with common genetic risk variants for CNS disorders by leveraging genome-wide association studies (GWAS) summary statistics of four major neuropsychiatric disorders and several neurological traits from the UK Biobank. We found moderate co-localization ($PPH4$ 0.3-0.8) for 98 loci across 15 traits and disorders and strong evidence of co-localization ($PPH4$ > 0.8) for 9 loci across 4 traits and disorders (**Supplemental Table 9, Figure 8F**). The majority of disease variants co-localized with postnatal predominant edQTLs ($n=72$) followed by consistent ($n=20$) and prenatal predominant edQTLs ($n=15$). Postnatal edQTLs uniquely co-localized with variants that implicate risk for mild cognitive impairment and dementia while prenatal edQTLs uniquely co-localized with variants that implicate risk for generalized seizures and hemiplegia (**Supplemental Table 9**). Both consistent and prenatal predominant edQTLs co-localized with sleeping disorder traits, including two editing sites in the third intron of *C16orf72* ($PPH4=0.99$, $PPH4=0.95$, respectively) (**Figure 8H-I**). Notably, we also observed significant co-localization, with similar direction of effect, for a risk variant for attention deficit hyperactivity disorder (ADHD) with a prenatal

predominant edQTL linked to an editing site in the 3'UTR of *PCDH7* (*PPH4*=0.95) (**Figure 8J**). The same edQTL also co-localized with a separate variant associated with schizophrenia (*PPH4*=0.57).

Discussion

This work highlights RNA editing as an important, and relatively underexplored, species-conserved mechanism involved in cortical development. We provide a global temporal picture of RNA editing in the developing human cortex, from early progenitor cells to mature cells in centenarians. We recapitulated RNA editing profiles in non-human primate and murine models of neurodevelopment, underscoring RNA editing as a conserved evolutionary regulatory mechanism in mammalian cortical development. While these data do not immediately lend themselves to quick mechanistic interpretation, at each level of analysis we provide tangible examples and proofs-of-principle that establish starting lines for investigations to link RNA editing with mechanisms of cortical development.

Our analysis is anchored around BrainVar data, permitting precise base-specific quantifications of RNA editing from RNA-seq data guided by paired WGS (**Figure 1**). We observed that the late-fetal transition in gene expression between mid-fetal development and infancy is synchronized with the peak of RNA editing, that involves over 10,000 RNA editing sites that increase in editing efficiency throughout cortical development (**Figure 3**). This represents a substantial departure from the current perception of RNA editing in brain development²⁷⁻³². Our results, akin to other recent reports^{33,34}, indicate that these developmental periods are marked by clear changes in the proportions and compositions of various cell types in the brain, including the distinct transcriptional and RNA editing programs within these cell subsets. Recognizing the significance of this, we accounted for the proportion of mature neurons for each of our analyses, which preserved the significance of our findings and together indicated an increased rate of editing per cell through development. Providing these editing sites are involved in processes related to neuronal differentiation, mRNA metabolism and binding of actin cytoskeleton proteins and 3'UTR elements (**Figure 4**), alterations or reversals of these editing rates in the context of neurodevelopmental disorders, like schizophrenia (**Figure 4H**), may provide avenues for understanding their potential pathological impact.

Likewise, we uncover numerous temporally regulated recoding sites enriched for RNA metabolism and binding of cadherin and glutamate as well as genes abundant around the postsynaptic membrane (**Figure 5**). These results confirm several previously described recoding sites known to regulate Ca^{2+} permeability (*GRIK2*, Y/C, ~58% increase)^{16,17}, remodel actin cytoskeletal at excitatory synapses (*CYFIP2*, K/E, ~50% increase)^{19,20} and guide gating kinetics of inhibitory receptors (*GABRA3*, I/M, ~56% increase)^{21,22}, but also highlight several sites with unexplored roles in neurodevelopment. For example, the Q/R site in *FLNA* increases ~27% through development, a gene that anchors membrane receptors to the actin cytoskeleton and regulates their precise location and transport. The R/G site in cyclin-I (*CCNI*), a gene characterized by dramatic periodicity in protein abundance through the cell cycle and important for the immune response³⁷, and the E/G site in calcium-dependent secretion activator (*CADPS*), previously shown to facilitate the rapid release of catecholamine and promote dense core vesicle exocytosis³⁸ both increase in editing efficiency by ~15%. However, the functional roles of these sites in cortical development have yet to be studied. We also identified seven novel recoding sites with high prenatal editing rates in regions of the genome with high mappability, which were not confounded by sequencing or other genome artifacts. These include a K/E site (chr8:73709034) in the double-stranded RNA binding motif of staufen double-stranded RNA binding protein 2 (*STAU2*). This gene is widely expressed in the brain, with high early fetal expression and is required for microtubule-dependent transport of neuronal RNA from the cell body to the dendrite³⁹. The D/G site (chr1:9930448) in the beta-catenin-interacting protein domain of leucine zipper and

CTNNB1P1 domain containing gene (*LZIC*), may play a role in early developmental binding and regulation of beta-catenin^{40,41}. A Q/R site (chr1:213071037) in the PX domain Ribosomal Protein S6 Kinase C1 (*RPS6KC1*), a phosphoinositide-binding structural domain involved in targeting of proteins to cell membranes⁴². Collectively, the recoding sites defined here have yet to be studied thoroughly in the context of neuronal maturation and cortical development and highlight numerous entry points for functional investigation.

RNA hyper-editing signal dramatically increased in frequency through development into advanced aging. Stretches of hyper-editing were predominantly found in Alu rich regions of the genome and more specifically, within introns and 3'UTRs that putatively form double-stranded RNA structures, consistent with recent reports⁴³⁻⁴⁵. Our analysis refined the developmental timing of these events and identified hundreds of specific genes that accumulate hyper-editing sites through development, most of which occur within retained introns of ion-channels and genes essential for typical development (**Figure 6**). Most hyper-edited regions are seldom detected as non-edited and thus may represent key ADAR targets. Markedly, hyper-editing events amassed in ultra-conserved non-coding segments, sequences that exhibit extremely high conservation across vertebrate genomes and cluster in non-coding regions of genes coding for transcription factors and key developmental regulators^{46,47}. Thus, build-up of age-related RNA hyper-editing in these regions, and corresponding gene products, might underlie their suboptimal regulation and activity during aging, contributing to senescent age-related phenotypes.

While ADAR enzymes catalyze A-to-I RNA editing, the degree to which common genetic variation sets the mode and tempo of editing levels through development is a novel line of enquiry. We identified hundreds of edQTLs that have a greater effect on editing levels prenatally or postnatally (**Figure 8, Supplemental Table 8-9**). When combined with findings from large-scale GWAS, consistent and temporal-predominant edQTLs co-localized with a total of several neurological traits and disorders, including sleep disorders, ADHD, schizophrenia and cognitive impairment. Given that most risk loci reside in non-coding regions, the vast majority of edQTLs and genetic colocalization occurred with RNA editing sites in 3'UTRs, and did occur in coding regions or with RNA recoding events. We identified ADHD risk variant rs28522755 (chr4: 31143818 A/G)⁴⁸ co-localized with a consistent edQTL linked to an A-to-G editing site in the 3'UTR of *PCDH7*. A separate postnatal predominant edQTL, linked to the same 3'UTR editing site in *PCDH7*, also co-localized with risk variant rs13145415 (chr4:30823157 T/G) for schizophrenia⁴⁹. *PCDH7* plays a significant role in axonal guidance, predominantly through neuronal cell adhesion and short range signaling, and it has been proposed that mutations in *PCDH7* could introduce alterations in neuronal localization and perhaps brain structure⁵⁰. Taken together, these two variants might exert their adverse consequences through altered 3'UTR RNA editing, shaping *PCDH7* gene expression and function, potentially offering new therapeutic avenues. Notwithstanding sample size limitations, these results identify temporal predominant edQTLs and GWAS-edQTL co-localization, thereby generating specific hypotheses for the role of RNA editing regulation in cortical development and neurological disease etiology.

In sum, RNA editing acts as conserved epitranscriptional mechanism in cortical physiology and is characterized by a global expansion of RNA sequence diversity. RNA editing rates increase throughout neurodevelopment, are commonly genetically regulated, and occur disproportionately in 3'UTRs of neurodevelopmental genes. This work further highlights variation in RNA editing as an intermediate mechanism that influences phenotypes of complex neurological traits and disorders. Collectively, these results serve as a resource toward classifying functionally relevant editing sites in the brain, as they uncover highly developmentally and genetically regulated sites which can serve to bridge the gap between their function and regulation of complex traits.

REFERENCES AND NOTES

1. Behm, M. & Öhman, M. RNA editing: a contributor to neuronal dynamics in the mammalian brain. *Trends Genet.* 32, 165–175 (2016).
2. Rosenthal, J. J. The emerging role of RNA editing in plasticity. *Journal of Experimental Biology*, 218(12), 1812-1821 (2015).
3. Bazak, L., et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome research*, 24(3), 365-376 (2014).
4. Carmi, S., Borukhov, I., & Levanon, E. Y. (2011). Identification of widespread ultra-edited human RNAs. *PLoS Genet*, 7(10), e1002317.
5. Paz-Yaacov, N., Levanon, E. Y., Nevo, E., Kinar, Y., Harmelin, A., Jacob-Hirsch, J., ... & Rechavi, G. (2010). Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proceedings of the National Academy of Sciences*, 107(27), 12174-12179.
6. Mehler, M. F., & Mattick, J. S. (2007). Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiological reviews*, 87(3), 799-823.
7. Tan, M. H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A. N., ... & Li, J. B. (2017). Dynamic landscape and regulation of RNA editing in mammals. *Nature*, 550(7675), 249-254.
8. Basilio, C., Wahba, A. J., Lengyel, P., Speyer, J. F., & Ochoa, S. (1962). Synthetic polynucleotides and the amino acid code, V. *Proceedings of the National Academy of Sciences of the United States of America*, 48(4), 613.
9. Wahlstedt, H., & f# x2013; hman, M. (2011). Site-selective versus promiscuous A-to-I editing. *Wiley Interdisciplinary Reviews: RNA*, 2(6), 761-771.
10. Walkley, C. R., & Li, J. B. (2017). Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. *Genome biology*, 18(1), 1-13.
11. Porath, H. T., Carmi, S., & Levanon, E. Y. (2014). A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nature communications*, 5(1), 1-10.
12. Xu, G., & Zhang, J. (2014). Human coding RNA editing is generally nonadaptive. *Proceedings of the National Academy of Sciences*, 111(10), 3769-3774.
13. Laurencikiene, J., Källman, A. M., Fong, N., Bentley, D. L., & Öhman, M. (2006). RNA editing and alternative splicing: the importance of co-transcriptional coordination. *EMBO reports*, 7(3), 303-307.
14. Park, E., Williams, B., Wold, B. J., & Mortazavi, A. (2012). RNA editing in the human ENCODE RNA-seq data. *Genome research*, 22(9), 1626-1633.
15. Yang, W., Chendrimada, T. P., Wang, Q., Higuchi, M., Seeburg, P. H., Shiekhattar, R., & Nishikura, K. (2006). Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nature structural & molecular biology*, 13(1), 13-21.
16. Sommer, B., Köhler, M., Sprengel, R., & Seeburg, P. H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell*, 67(1), 11-19.
17. Burnashev, N., Monyer, H., Seeburg, P. H., & Sakmann, B. (1992). Divalent ion permeability of AMPA receptor channels is dominated by the edited form of a single subunit. *Neuron*, 8(1), 189-198.
18. Lomeli, H., Mosbacher, J., Melcher, T., Hoger, T., Kuner, T., Monyer, H., ... & Seeburg, P. H. (1994). Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science*, 266(5191), 1709-1713.
19. Bonini, D., Filippini, A., La Via, L., Fiorentini, C., Fumagalli, F., Colombi, M., & Barbon, A. (2015). Chronic glutamate treatment selectively modulates AMPA RNA editing and ADAR expression and activity in primary cortical neurons. *RNA biology*, 12(1), 43-53.
20. Nishimoto, Y., Yamashita, T., Hideyama, T., Tsuji, S., Suzuki, N., & Kwak, S. (2008). Determination of editors at the novel A-to-I editing positions. *Neuroscience research*, 61(2), 201-206.
21. Ohlson, J., Pedersen, J. S., Haussler, D., & Öhman, M. (2007). Editing modifies the GABA_A receptor subunit $\alpha 3$. *Rna*, 13(5), 698-703.

22. Daniel, C., Wahlstedt, H., Ohlson, J., Björk, P., & Öhman, M. (2011). Adenosine-to-inosine RNA editing affects trafficking of the γ -aminobutyric acid type A (GABA A) receptor. *Journal of Biological Chemistry*, 286(3), 2031-2040.

23. Ohlson, J., Ensterö, M., Sjöberg, B. M., & Öhman, M. (2005). A method to find tissue-specific novel sites of selective adenosine deamination. *Nucleic acids research*, 33(19), e167-e167.

24. Licht, K., Kapoor, U., Amman, F., Picardi, E., Martin, D., Bajad, P., & Jantsch, M. F. (2019). A high resolution A-to-I editing map in the mouse identifies editing events controlled by pre-mRNA splicing. *Genome research*, 29(9), 1453-1463.

25. Lundin, E., Wu, C., Widmark, A., Behm, M., Hjerling-Leffler, J., Daniel, C., ... & 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 biology*, 18(1), 1-15.

26. Wahlstedt, H., Daniel, C., Ensterö, M., & Öhman, M. (2009). Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome research*, 19(6), 978-986.

27. Ekdahl, Y., Farahani, H. S., Behm, M., Lagergren, J., & Öhman, M. (2012). A-to-I editing of microRNAs in the mammalian brain increases during development. *Genome research*, 22(8), 1477-1487.

28. Shtrichman R, Germanguz I, Mandel R, Ziskind A, Nahor I, Safran M, Osenberg S, Sherf O, Rechavi G, Itskovitz-Eldor J. Altered A-to-I RNA editing in human embryogenesis. *PLoS One*. 2012 Jul 31;7(7):e41576.

29. Li Z, Bammann H, Li M, Liang H, Yan Z, Chen YP, Zhao M, Khaitovich P. Evolutionary and ontogenetic changes in RNA editing in human, chimpanzee, and macaque brains. *Rna*. 2013 Dec 1;19(12):1693-702.

30. Nicholas, A., De Magalhaes, J. P., Kraytsberg, Y., Richfield, E. K., Levanon, E. Y., & Khrapko, K. (2010). Age-related gene-specific changes of A-to-I mRNA editing in the human brain. *Mechanisms of ageing and development*, 131(6), 445-447.

31. Hwang, T., Park, C. K., Leung, A. K., Gao, Y., Hyde, T. M., Kleinman, J. E., ... & Weinberger, D. R. (2016). Dynamic regulation of RNA editing in human brain development and disease. *Nature neuroscience*, 19(8), 1093-1099.

32. Gordon, A., Yoon, S.J., Tran, S.S., Makinson, C.D., Park, J.Y., Andersen, J., Valencia, A.M., Horvath, S., Xiao, X., Huguenard, J.R. and Paşa, S.P., (2021). Long-term maturation of human cortical organoids matches key early postnatal transitions. *Nature Neuroscience*, 24(3), pp.331-342.

33. Werling, D. M., Pochareddy, S., Choi, J., An, J. Y., Sheppard, B., Peng, M., ... & Sestan, N. (2020). Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex. *Cell reports*, 31(1), 107489.

34. Li, M., Santpere, G., Kawasawa, Y. I., Evgrafov, O. V., Gulden, F. O., Pochareddy, S., ... & Sestan, N. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*, 362(6420).

35. Jain, M., Mann, T. D., Stulić, M., Rao, S. P., Kirsch, A., Pullirsch, D., ... & Jantsch, M. F. (2018). RNA editing of Filamin A pre-mRNA regulates vascular contraction and diastolic blood pressure. *The EMBO journal*, 37(19), e94813.

36. Czermak, P., Amman, F., Jantsch, M. F., & Cimatti, L. (2018). Organ-wide profiling in mouse reveals high editing levels of Filamin B mRNA in the musculoskeletal system. *RNA biology*, 15(7), 877-885.

37. Zhang, M., Fritsche, J., Roszik, J., Williams, L. J., Peng, X., Chiu, Y., ... & Hwu, P. (2018). RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nature communications*, 9(1), 1-10.

38. Miyake, K., Ohta, T., Nakayama, H., Doe, N., Terao, Y., Oiki, E., ... & Kawahara, Y. (2016). CAPS1 RNA editing promotes dense core vesicle exocytosis. *Cell reports*, 17(8), 2004-2014.

39. Goetze, B., Tuebing, F., Xie, Y., Dorostkar, M. M., Thomas, S., Pehl, U., ... & Kiebler, M. A. (2006). The brain-specific double-stranded RNA-binding protein Staufen2 is required for dendritic spine morphogenesis. *The Journal of cell biology*, 172(2), 221-231.

40. Clements, W. K., & Kimelman, D. (2005). LZIC regulates neuronal survival during zebrafish development. *Developmental biology*, 283(2), 322-334.

41. Katoh, M. (2001). Molecular cloning and characterization of LZIC, a novel gene encoding ICAT homologous protein with leucine zipper domain. *International journal of molecular medicine*, 8(6), 611-615.
42. Liu, L., Yang, C., Yuan, J., Chen, X., Xu, J., Wei, Y., ... & Yu, L. (2005). RPK118, a PX domain-containing protein, interacts with peroxiredoxin-3 through pseudo-kinase domains. *Molecules & Cells (Springer Science & Business Media BV)*, 19(1).
43. Porath, H. T., Knisbacher, B. A., Eisenberg, E., & Levanon, E. Y. (2017). Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. *Genome biology*, 18(1), 1-12.
44. Porath, H. T., Carmi, S., & Levanon, E. Y. (2014). A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nature communications*, 5(1), 1-10.
45. Roth, S. H., Danan-Gotthold, M., Ben-Izhak, M., Rechavi, G., Cohen, C. J., Louzoun, Y., & Levanon, E. Y. (2018). Increased RNA editing may provide a source for autoantigens in systemic lupus erythematosus. *Cell reports*, 23(1), 50-57.
46. Dimitrieva, S., & Bucher, P. (2013). UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic acids research*, 41(D1), D101-D109.
47. Prabhakar, S., Noonan, J. P., Pääbo, S., & Rubin, E. M. (2006). Accelerated evolution of conserved noncoding sequences in humans. *science*, 314(5800), 786-786.
48. Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., ... & Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature genetics*, 51(1), 63-75.
49. Ripke, S., Walters, J. T., O'Donovan, M. C., & Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*.
50. Dark, C., Homman-Ludiye, J., & Bryson-Richardson, R. J. (2018). The role of ADHD associated genes in neurodevelopment. *Developmental biology*, 438(2), 69-83.
51. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21.
52. Flati, T., Gioiosa, S., Spallanzani, N., Tagliaferri, I., Diroma, M. A., Pesole, G., ... & Castrignanò, T. (2020). HPC-REDItools: a novel HPC-aware tool for improved large scale RNA-editing analysis. *BMC bioinformatics*, 21(10), 1-12.
53. Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38(16), e164-e164.
54. Picardi, E., D'Erchia, A. M., Lo Giudice, C., & Pesole, G. (2017). REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic acids research*, 45(D1), D750-D757.
55. Hubisz, M. J., Pollard, K. S., & Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Briefings in bioinformatics*, 12(1), 41-51.
56. Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE blacklist: identification of problematic regions of the genome. *Scientific reports*, 9(1), 1-5.
57. Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, 1-68.
58. Breen, M. S., Dobbyn, A., Li, Q., Roussos, P., Hoffman, G. E., Stahl, E., ... & Buxbaum, J. D. (2019). Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nature neuroscience*, 22(9), 1402-1412.
59. Roth, S. H., Levanon, E. Y., & Eisenberg, E. (2019). Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nature methods*, 16(11), 1131-1138.
60. Wang, J., Devlin, B., & Roeder, K. (2020). Using multiple measurements of tissue to estimate subject- and cell-type-specific gene expression. *Bioinformatics*, 36(3), 782-788.
61. Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., ... & Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences*, 112(23), 7285-7290.

62. Ritchie, M. E., Phipson, B., Wu, D. I., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, 43(7), e47-e47.

63. Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9(1), 1-13.

64. Chen, J., Bardes, E. E., Aronow, B. J., & Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research*, 37(suppl_2), W305-W311.

65. Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS one*, 6(7), e21800.

66. Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J. Y., ... & Demontis, D. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, 180(3), 568-584.

67. Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., ... & O'Donovan, M. C. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature*, 506(7487), 179-184.

68. Parikshak, N. N., Luo, R., Zhang, A., Won, H., Lowe, J. K., Chandran, V., ... & Geschwind, D. H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*, 155(5), 1008-1021.

69. Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., ... & Børglum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics*, 51(3), 431-444.

70. Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, 50(8), 1112-1121.

71. Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), 421-427.

72. Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., ... & Lichtner, P. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics*, 46(11), 1173-1186.

73. Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., ... & Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, 474(7351), 380-384.

74. Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., ... & PsychENCODE Consortium. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*, 362(6420).

75. Gandal, M. J., Haney, J. R., Parikshak, N. N., Leppa, V., Ramaswami, G., Hartl, C., ... & iPSYCH-BROAD Working Group. (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*, 359(6376), 693-697.

76. Tran, S. S., Jun, H. I., Bahn, J. H., Azghadi, A., Ramaswami, G., Van Nostrand, E. L., ... & Xiao, X. (2019). Widespread RNA editing dysregulation in brains from autistic individuals. *Nature neuroscience*, 22(1), 25-36.

77. Shen, L. (2014). GeneOverlap: An R package to test and visualize gene overlaps. *R Package*.

78. Betel, D., Koppal, A., Agius, P., Sander, C., & Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome biology*, 11(8), 1-14.

79. Jaganathan, K., Panagiotopoulou, S. K., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., ... & Farh, K. K. H. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, 176(3), 535-548.

80. Yeom, K. H., Pan, Z., Lin, C. H., Lim, H. Y., Xiao, W., Xing, Y., & Black, D. L. (2021). Tracking pre-mRNA maturation across subcellular compartments identifies developmental gene regulation through intron retention and nuclear anchoring. *Genome Research*, 31(6), 1106-1119.

81. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T., & Delaneau, O. (2016). Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*, 32(10), 1479-1485.

82. Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., ... & Lee, S. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature genetics*, 50(9), 1335-1341.

83. Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., ... & Reif, A. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics*, 51(5), 793-803.

84. Hyde, C. L., Nagle, M. W., Tian, C., Chen, X., Paciga, S. A., Wendland, J. R., ... & Winslow, A. R. (2016). Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics*, 48(9), 1031-1036.

85. Dobbyn, A., Huckins, L. M., Boocock, J., Sloofman, L. G., Glicksberg, B. S., Giambartolomei, C., ... & Klei, L. L. (2018). Landscape of conditional eQTL in dorsolateral prefrontal cortex and co-localization with schizophrenia GWAS. *The American Journal of Human Genetics*, 102(6), 1169-1184.

ACKNOWLEDGEMENTS

Funding: MSB is a Seaver Foundation Faculty Scholar. U01MH122678 (NS and SS), U01AG046170 (BZ), U01AG052411 (BZ), RF1AG057440 (BZ).

Author contributions:

Conceptualization: MSB

Data analysis: MSB, RC, EM, LS, LL, XF, MW, JW

Visualization: MSB, RC

Funding acquisition: MSB, SS, BZ, BD, KR, NS, JB

Supervision: MSB, BZ, BD, KR, JB, NS, SS

Writing – original draft: MSB, RC, LS, LL, XF, MW, BZ, JW, BD, KR, JB, SS

Competing interests: Authors declare that they have no competing interests

Data and materials availability: The majority of data are available in supplementary tables. All code and larger data tables and files are available at GitHub (<https://github.com/BreenMS>).

Main Figure Legends

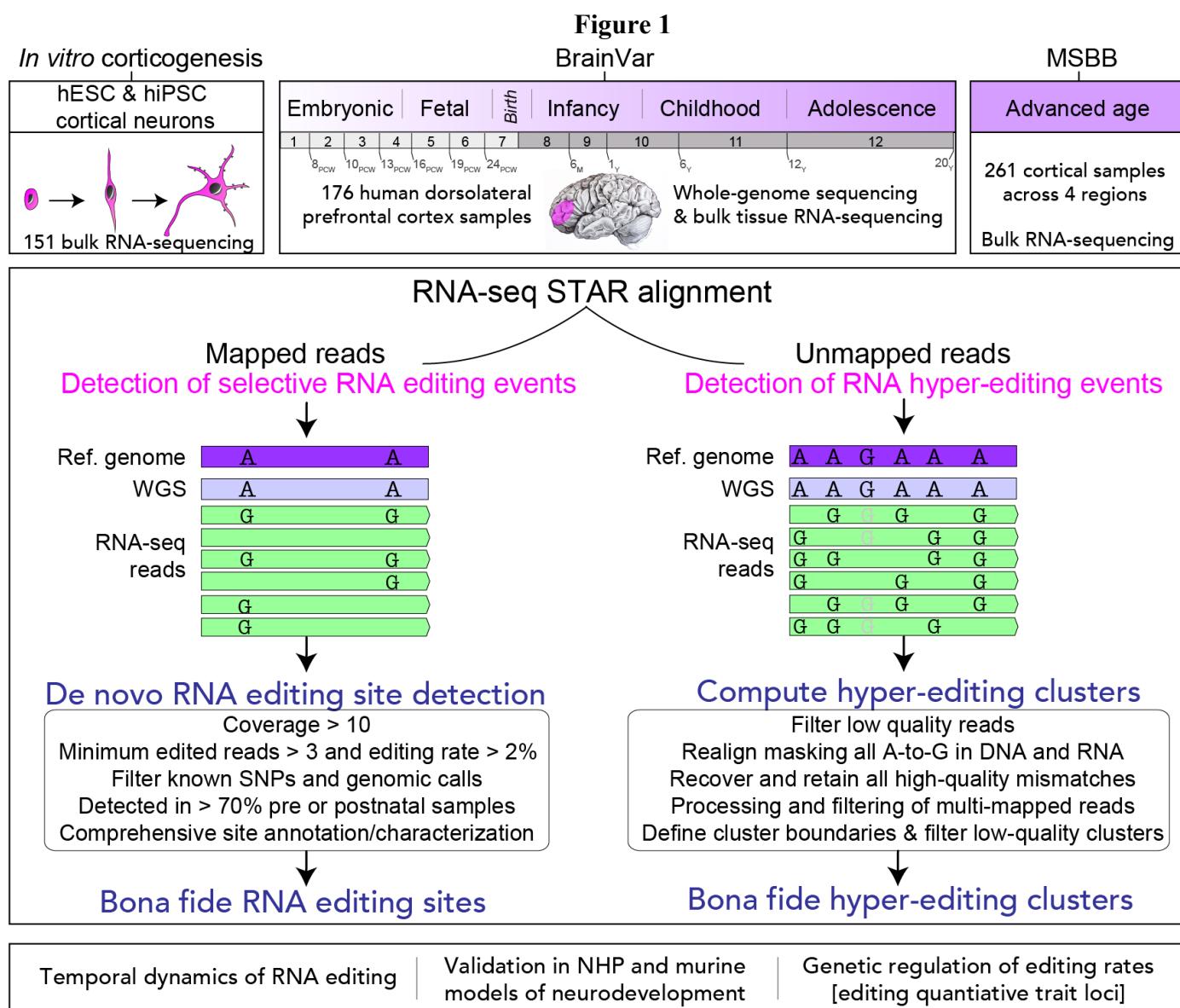


Figure 1. A general overview of the current study. Our analysis is anchored around paired WGS and RNA-seq data from BrainVar ($n=176$, DFLPC), together with *in vitro* human embryonic stem cell (hESC, $n=24$) and human pluripotent stem cell (hiPSC, $n=127$) models of corticogenesis and cortical samples from advanced stages of aging ($n=261$, 61-108 years). We also provide validation in a smaller developmental bulk brain transcriptome dataset from the forebrain ($n=55$) and hindbrain ($n=59$). We evaluate both selective RNA editing events from mapped RNA-seq bam files and RNA hyper-editing events from unmapped RNA-seq fastq files (see Methods). Our analytical approach sought to identify developmental trends in editing rates, provide evolutionary context in animal models of neurodevelopment, and elucidate common genetic variants that exert their effects by altering editing levels.

Figure 2

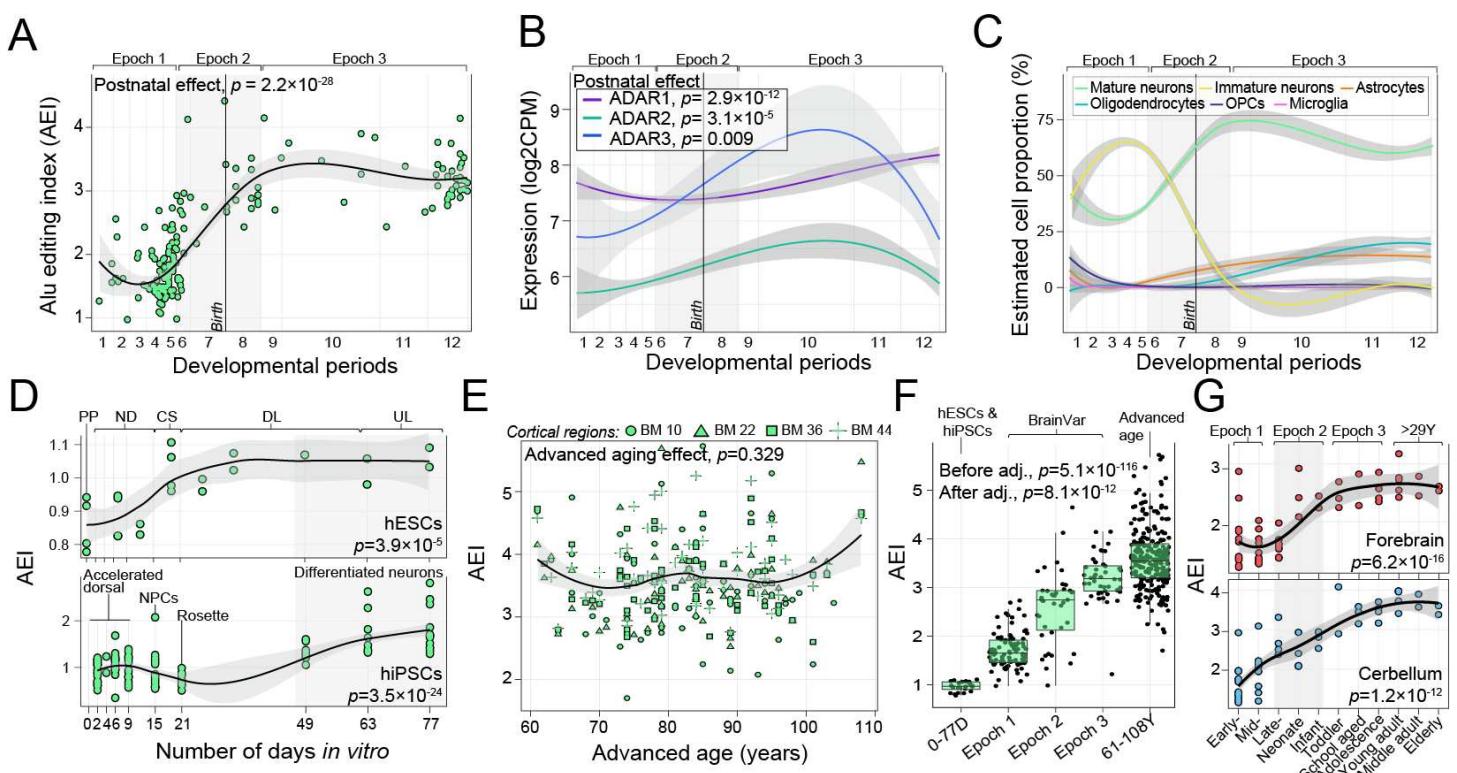


Figure 2. Alu editing index through cortical development. (A) The Alu editing index (AEI; y-axis), (B) expression profiles for *ADAR1*, *ADAR2* and *ADAR3* (y-axis) and (C) estimated cell type proportions based on cell-specific signatures from Darmanis et al., 2015 (y-axis) were computed for each DLPFC sample in BrainVar and examined across 12 developmental periods (log age, x-axis). Periods 1-7 reflect prenatal windows and periods 8-12 reflect postnatal windows. A linear regression was used to test for significance. (D) The AEI (y-axis) in human embryonic stem cell (hESC) and human induced pluripotent stem cell (hiPSC) models of corticogenesis across 77 *in vitro* (x-axis). (E) The AEI (y-axis) across advanced ages in the MSBB cohort (61-108 years). (x-axis). (H) The AEI (y-axis) across early hESCs, BrainVar (1st, 2nd and 3rd Epochs), and into advanced aging (MSBB). A linear regression was used to test for significance both with and without adjustment for neuronal cell type proportions. (G) Validation of the AEI in independent bulk forebrain ($n=55$) and cerebellar ($n=59$) bulk tissue transcriptome data across the lifespan.

Figure 3

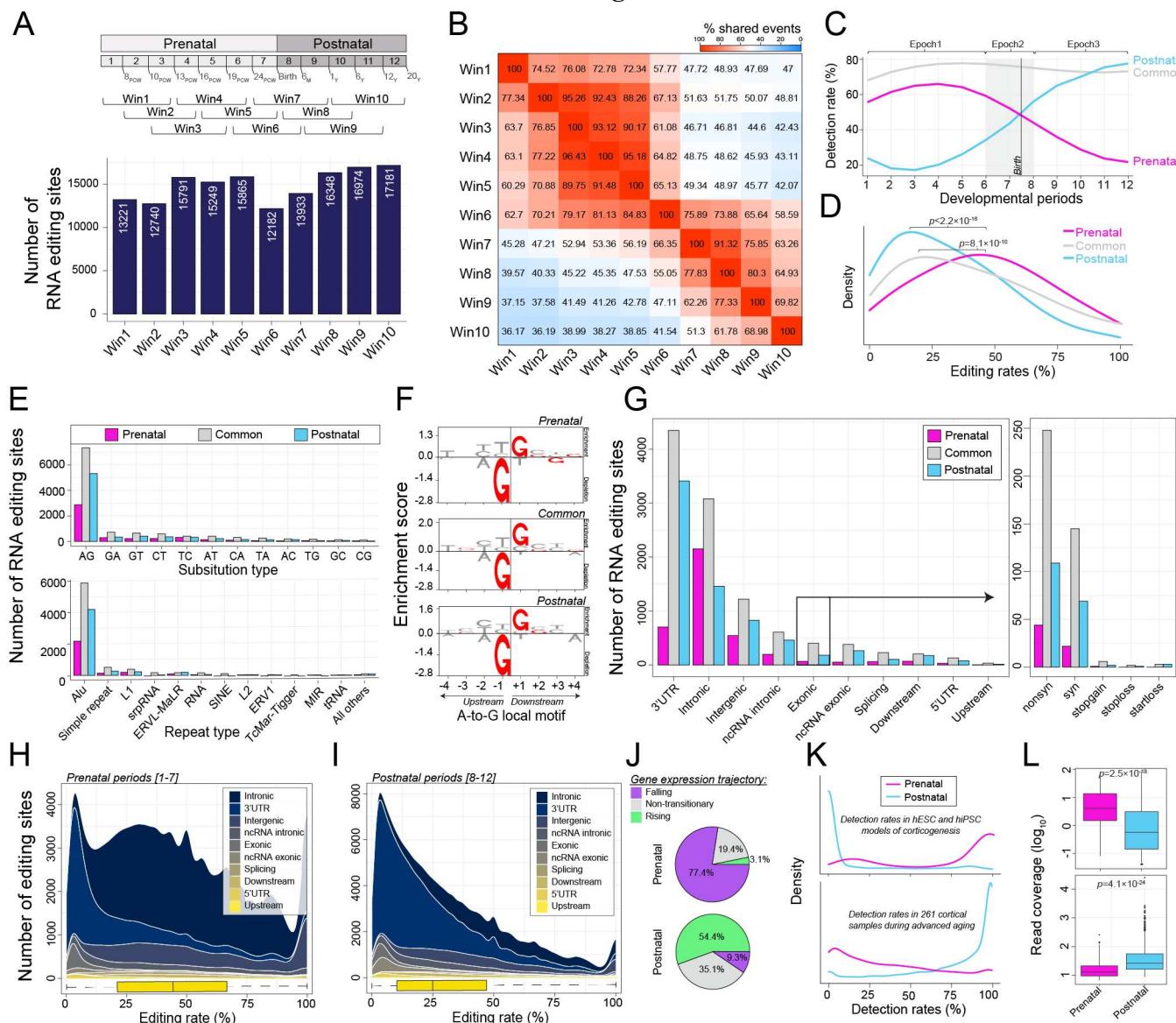


Figure 3. Identification and annotation of selective RNA editing sites. (A) A sliding window approach identified sites shared across 80% of samples per window, and a uniform distribution of editing sites were detected across all windows. (B) Sites specific to each window where then overlapped onto each other to reveal an enrichment of prenatal-specific (N sites=3,879) and postnatal-specific (N sites=6,969) sites, as well as sites found across all periods, termed ‘common sites’ (N sites =10,652). (C) Detection rates (%) indicate the fraction of samples for which prenatal, postnatal and common sites are detected with sufficient coverage per period. (D) Distribution of RNA editing rates (%) (x-axis) across all prenatal, postnatal and common sites. Significance was computed using a Mann-Whitney U test. (E) Detected sites were annotated according to substitution type (upper) and to repeat elements (lower). (F) A-to-G sites enriched (y-axis) for a conserved local sequence motif, featuring a depletion of guanosine +1bp upstream and enrichment of guanosine -1bp downstream the target adenosine. (G) Sites according to genic region and an enrichment of 3'UTR sites among postnatal-specific sites. Distribution of editing rates among (H) prenatal-specific and (I) postnatal-specific sites by genic region. Here the amount of area reflects a larger contribution of editing sites from the corresponding genic region. (J) Gene expression trajectories for prenatal-specific sites largely decrease through development and trajectories for postnatal-specific sites

largely increase through development. (**K**) Detection rates (%, x-axis) of prenatal- and postnatal-specific sites in *in vitro* models of corticogenesis (upper panel) and 261 cortical samples of advanced age (lower panel) illustrate a high detection rate (**L**) and high read coverage (log₁₀ scale) of these corresponding temporal editing sites. Mann-Whitney U test was used to test for differences in coverage rates.

Figure 4

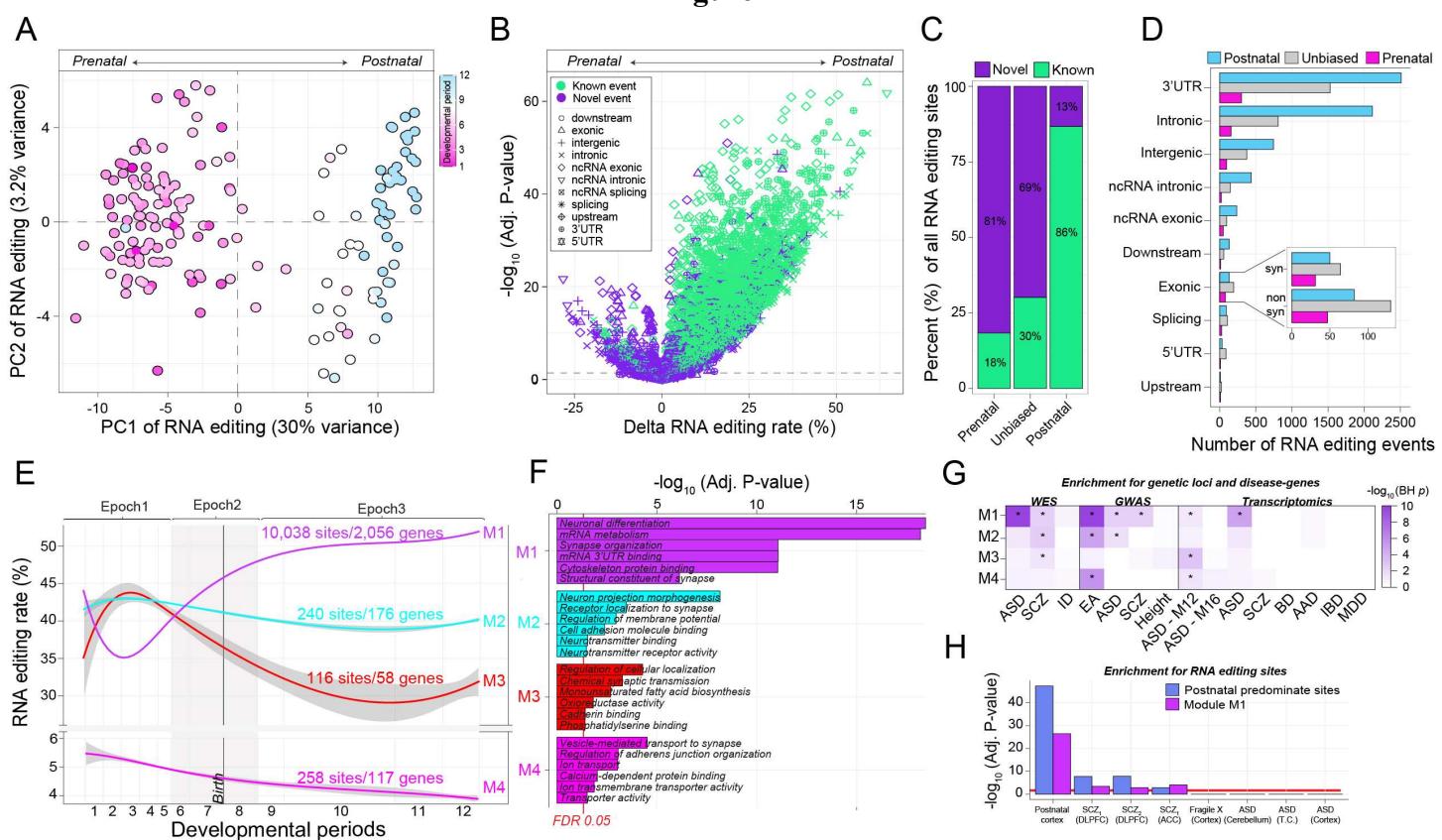


Figure 4. Developmental regulation of selective RNA editing sites. (A) Principal component analysis of RNA editing sites (N=10,652 sites) across all samples stratifies prenatal (pink) from postnatal (blue) periods. (B) Differential RNA editing analysis compares the strength of significance (-log₁₀ FDR-adjusted P; y-axis) of developmentally regulated sites relative to delta editing rates (x-axis). Sites according to the developmental bias (prenatal, postnatal or unbiased) are partitioned according to (C) their novelty defined as presence/absence from existing RNA editing databases and (D) by genic region. (E) Weighted correlation network analysis partitioned all sites into four modules of sites with coordinated editing patterns across development. The average editing rates and standard error across all sites within each module (y-axis) are plotted across cortical development (log age, x-axis). (F) Each site was mapped to a gene and tested for functional annotation. The top 6 significant functional GO terms are displayed for each module. (G) Similarly, each module was examined for enrichment of neurodevelopmental disorder-related genes and gene sets identified from large-scale genetic and genomic studies. Color intensity reflects strength of significance (-log₁₀ Adj. P). (H) Module M1 and all postnatal predominant sites were examined for enrichment of RNA editing sites found to be dysregulated in studies of postmortem brain tissue. Above the red line indicates a BH adjusted $p < 0.05$.

Figure 5

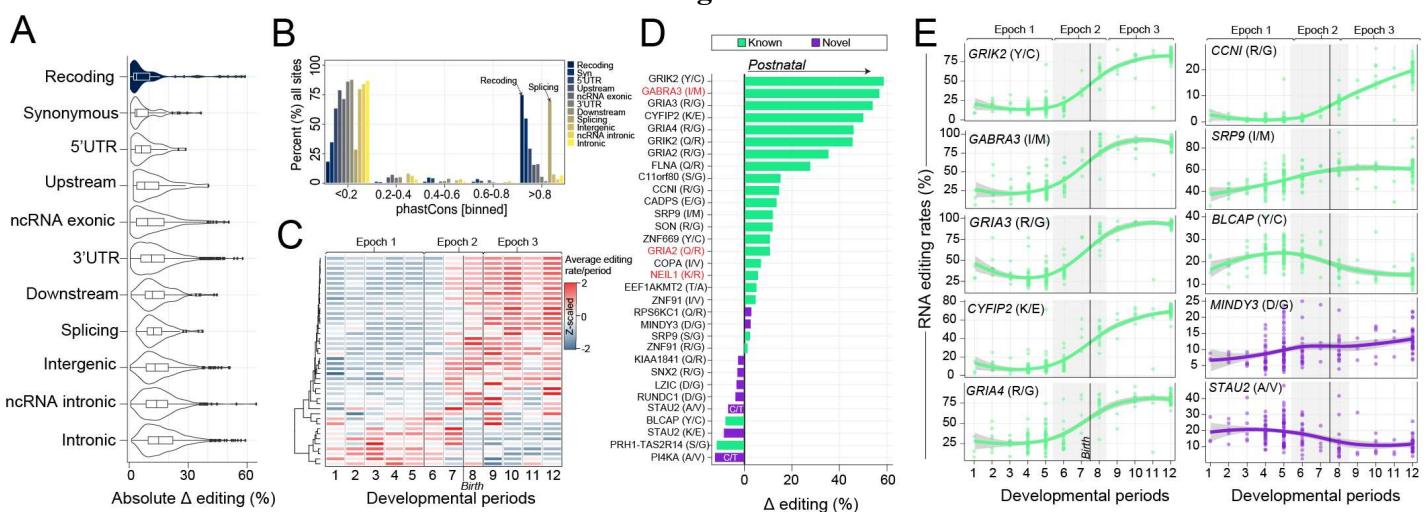


Figure 5. Temporal specificity of RNA recoding events. (A) Absolute delta editing rates parsed and ranked by genic region, where recoding events show the weakest developmental effect. (B) phastCons conservation metrics binned (x-axis) and the percent of all RNA editing sites per region are displayed per bin (y-axis). (C) Unsupervised clustering and heatmap depiction of 32 RNA editing recoding events that significantly change across development. (D) Ranking of 32 recoding sites by strength of developmental effect on editing rates (x-axis). Genes and sites in red indicate those where editing rates reach 100% and those in purple reflect novel recoding events. (E) Examples of editing rates (y-axis) for ten recoding sites across development, including five well-known sites (left), three known documented sites with unexplored roles in development (right) and two novel recoding (right) (x-axes). Loess curves were used to fit the data.

Figure 6

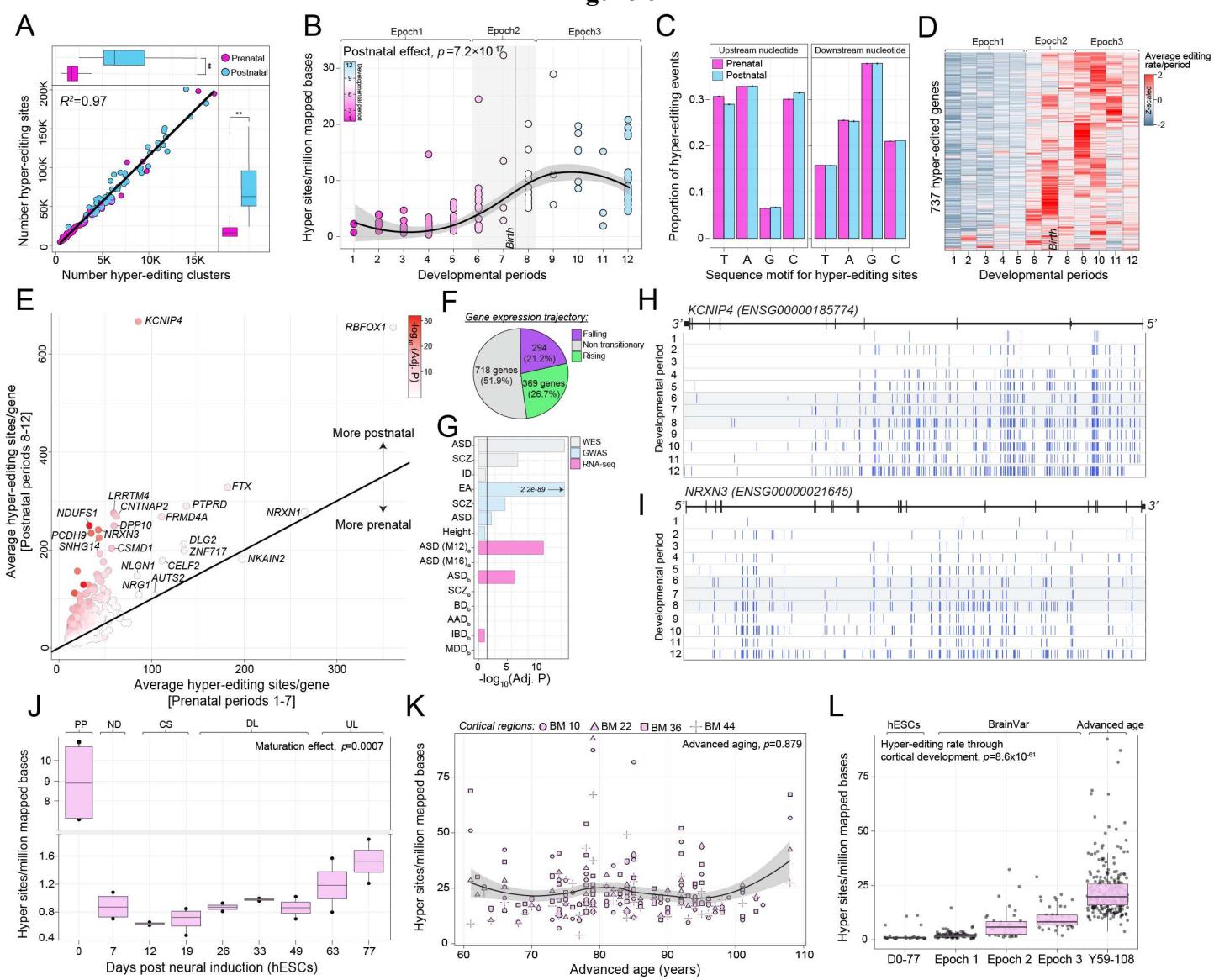


Figure 6. RNA hyper-editing signal through cortical development. (A) The number of hyper-editing sites (y-axis) is concordant with the number of hyper-editing clusters (x-axis). Significant increase ($**$, $p < 2 \times 10^{-16}$) in the number of hyper-editing sites and clusters was observed between prenatal and postnatal periods (linear regression). (B) The number of RNA hyper-editing sites normalized by the number of uniquely mapped bases per sample. This normalized measure of hyper-editing (y-axis) was examined across 12 developmental periods (x-axis) prior-to and following adjustment for the proportions of mature neurons with linear regression ($p = 7.2 \times 10^{-17}$, $p = 1.0 \times 10^{-11}$, respectively). (C) A local sequence motif for all hyper-editing events detected prenatally and postnatally are enriched for a depletion of guanosine -1bp upstream and an enrichment of guanosine +1bp downstream the target adenosine. (D) Heatmap depiction of genes that amass hyper-editing events into postnatal development. The number of hyper-editing events per period were averaged for each gene and z-scaled. (E) Average number of hyper-editing events per gene during prenatal periods 1-7 (x-axis) versus postnatal periods 8-12 (y-axis). Each gene is colored according to their FDR adjusted p -value significance. (F) The developmental expression trajectories for genes that amass A-to-G hyper-editing sites during postnatal periods. (G) Genes enriched for postnatal hyper-editing enrich for neurodevelopmental disorder associated genes and gene-sets curated from large-scale genomic studies (Fisher's exact test). RNA hyper-editing barcode plots illustrate when

and where hyper-editing events amass in two educational attainment genes: **(H)** *KCNIP4* (ENSG00000185774) and **(I)** *NRXN3* (ENSG0000021645). Normalized hyper-editing rates are examined prior-to and following adjusting for neuronal proportions across **(J)** a hESC model of corticogenesis ($p=0.0007$, $p=0.03$, respectively), **(K)** advanced aging ($p=0.8$, $p=0.2$, respectively) and **(L)** integrating hECS, BrainVar, and advanced aging ($p=p=8.6\times10^{-61}$, $p=0.004$, respectively). Note, cells during pluripotency (PP) were not included in these analysis as they represent extreme hyper-editing outliers.

Figure 7

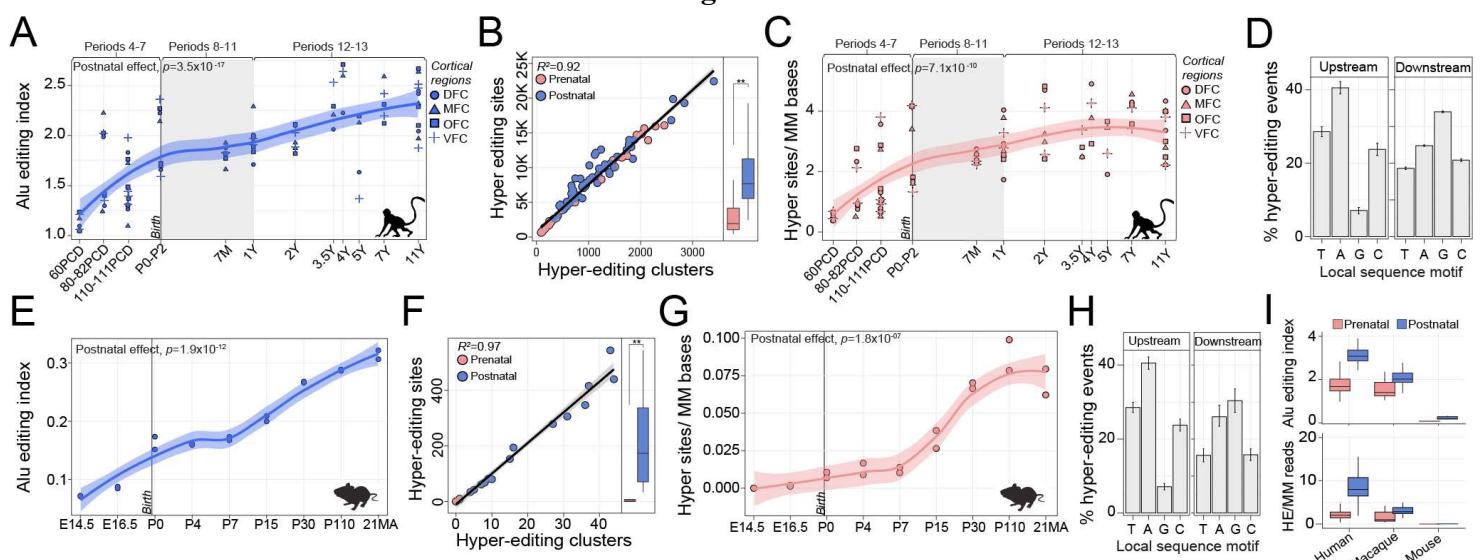


Figure 7. Evolutionary conservation of global editing in animal models of cortical development. (A) The AEI (y-axis) was computed for four cortical regions (DFC, MFC, OFC, VFC) in rhesus macaque across prenatal and postnatal periods (x-axis). Macaque developmental periods were matched with those closest to human (shown above) as previously shown (Y Zhu et al., 2018). (B) The number of hyper-editing sites (y-axis) is concordant with the number of hyper-editing clusters (x-axis) and the number of hyper-editing sites increases from prenatal to postnatal periods (**, indicates $p=2.8 \times 10^{-6}$). (C) Normalized hyper-editing rates across all developmental periods and cortical areas ($p=7.1 \times 10^{-7}$). (D) A local sequence motif for all hyper-editing events detected prenatally and postnatally are enriched for a depletion of guanosine -1bp upstream and an enrichment of guanosine +1bp downstream the target adenosine. (E) The AEI (y-axis) for whole cortex in mouse across nine developmental periods (x-axis). (F) The number of hyper-editing sites (y-axis) is concordant with the number of hyper-editing clusters (x-axis). Significant increase in the number of hyper-editing sites was observed between prenatal and postnatal periods (**, indicates $p<2 \times 10^{-16}$). (G) Normalized hyper-editing rates across all developmental periods. (H) A local sequence motif for all hyper-editing events detected are enriched for a depletion of guanosine -1bp upstream and an enrichment of guanosine +1bp downstream the target adenosine. (I) Comparison of the AEI and normalized hyper-editing rates across prenatal and postnatal stages between humans (using BrainVar), rhesus macaque and mouse. Linear regression was used to compute significance in all aforementioned tests.

Figure 8

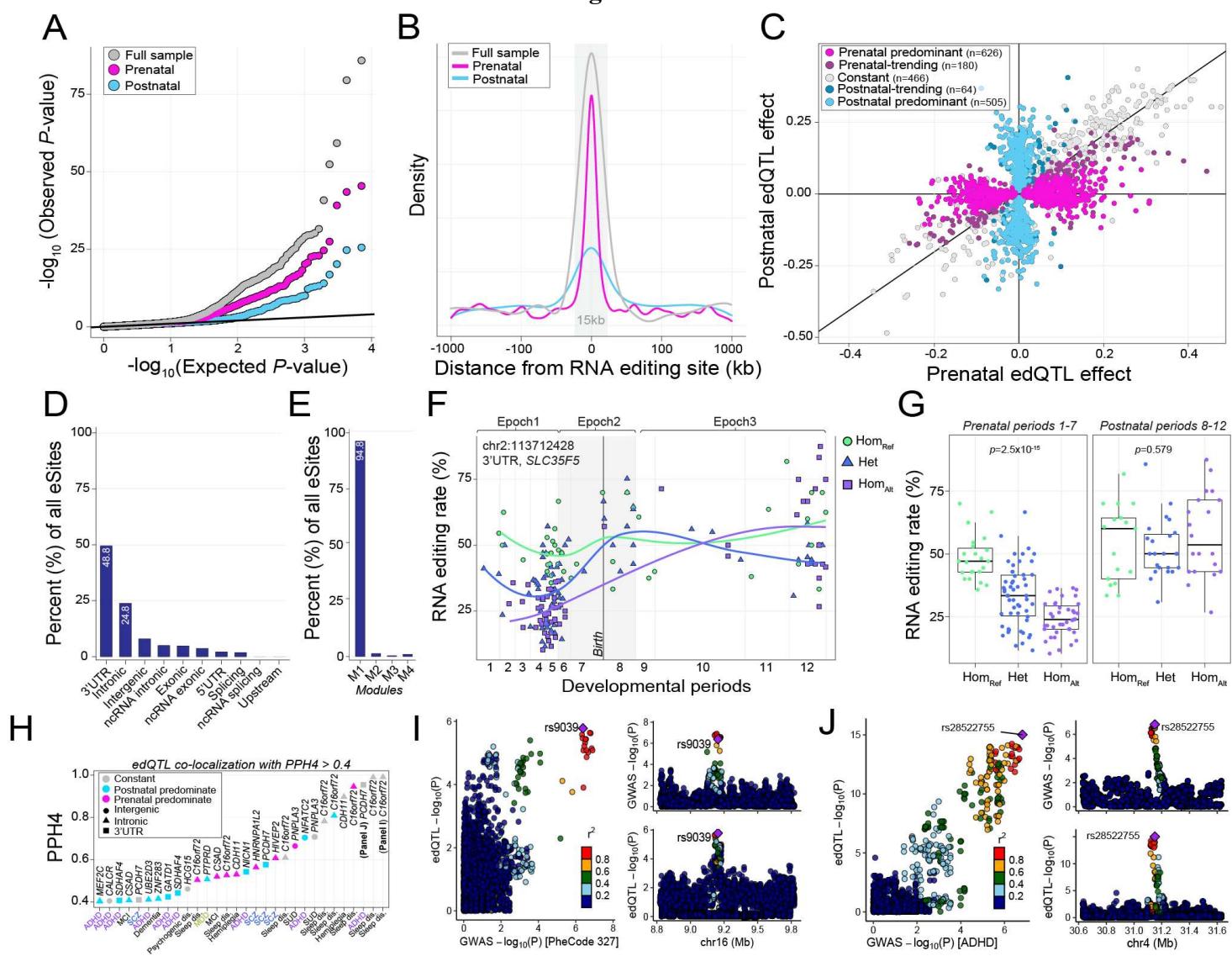


Figure 8. Cis-genetic regulation of RNA editing and co-localization with disease risk loci. (A) Quantile-quantile plot for association testing genome-wide P values between WGS-derived genotype dosages and 10,652 RNA editing sites across the full sample ($N=176$), prenatal periods 1-7 ($N=116$) and postnatal periods 8-12 ($N=60$) (linear regression and FDR correction via fastQTL). (B) Distribution of the association tests in relation to the distance between the editing site and variant for max-edQTLs. The grey box indicates ± 15 kb relative to the editing site. (C) Prenatal (x-axis) and postnatal (y-axis) effects for the edQTLs with the smallest p value for all eSites (points). The edQTLs are split into five categories based on temporal predominance using effect size and statistical thresholds; categories are represented by color. The percentages (y-axis) out of all eSites according to (D) genic regions and (E) modules (M) of coordinated editing sites. (F) RNA editing rates binned by genotype for a top prenatal-predominant edQTL for *SLC35F5*. Lines represent loess trajectories for RNA editing in samples with each of three genotypes. (G) Boxplots for prenatal (left) and postnatal (right) periods with each of three genotypes. (H) Summary of co-localization results for edQTLs with a posterior probability of H4 (PPH4; y-axis) greater than 0.4. Colors reflect temporal predominance of the edQTL effect and shapes indicate the RNA editing site genic region across all UK Biobank neurological traits (in black) and large-scale GWAS (colored). Locuscompare plots of the top co-localized hit for (I) variant rs9039 associated with sleep disorders (PheCode

327) and intronic A-to-G editing site in *C16orf72* (chr16:9110931) and (J) variant rs28522755 associated with ADHD and 3'UTR A-to-G editing site in *PCHD7* (chr4:31143136).