TITLE

Cell type and cortex-specific RNA editing in single human neurons informs neuropsychiatric disorders

Brendan Robert E. Ansell[1], Simon N. Thomas[1], Jacob E. Munro[1], Saskia Freytag[2], Melanie Bahlo[1]

1. *Population Health and Immunity Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia*
2. *Molecular Medicine Division, Harry Perkins Institute of Medical Research, Murdoch, Western Australia, Australia*

ABSTRACT

Conversion of adenosine to inosine in RNA by ADAR enzymes occurs at thousands of sites in the human transcriptome, and is essential for healthy brain development. This process, known as 'RNA editing', is dysregulated in many neuropsychiatric diseases, but is little understood at the level of individual neurons. We examined full-length nuclear transcriptomes of 3,055 neurons from six cortical regions of a neurotypical post-mortem female donor and identified 40,861 high-confidence edited sites. The majority of sites were located within Alu repeats in introns or 3' UTRs, and were present in previously published RNA editing databases. We identified 15,784 putative novel RNA editing sites, 30% of which were also detectable in independently generated neuronal transcriptomes from unrelated donors. The strongest correlates of global editing rates were expression levels of small nucleolar RNAs from the SNORD115 and SNORD116 cluster (15q11), known to modulate serotonin receptor processing and to colocalize with ADAR2, one of three known RNA editing enzymes in humans. As expected, expression of DNA and RNA binding proteins were negatively associated with editing. We present evidence for dysregulated RNA editing in six rare genetic conditions; and report 117 differentially edited sites between cortical regions and neuronal subtypes. These results provide spatial and neurophenotypic context for 1,871 and 998 sites that are differentially edited in the brains of schizophrenic and autistic patients respectively, and a reference for future studies of RNA editing in single brain cells from these cohorts.

INTRODUCTION

The extraordinary structural and functional complexity of the human brain arises via multiple layers of genetic regulation. The conversion of adenosine to inosine (A>I) in nascent RNA transcripts by ADAR1 and ADAR2 enzymes, known as 'RNA editing', is the most abundant RNA modification in the primate central nervous system, and confers transcriptomic diversity beyond that encoded in the genome [1]. RNA editing is essential for healthy brain development and increases with age [2].

Dysregulated editing is implicated in epilepsy [3], glioblastoma [4], major depression [5], autism spectrum disorder [6] and schizophrenia [7]. ADAR1 primarily edits adenosine within repetitive regions; ADAR2 primarily edits non-repetitive regions, and ADAR3 is a catalytically inactive inhibitor of editing [8]. ADAR2-null mice die *in utero* and partial knock-out neonatal animals succumb to severe seizures [9]. In humans, mutations in ADAR1 cause skin dyschromatosis (MIM 127400) and Aicardi-Goutieres encephalopathy (MIM 615010) with sub-types including striatal and motor neurodegeneration [10,11]. Mislocalization of ADAR was recently discovered in human and mouse models of C9orf72-mediated amyotrophic lateral sclerosis (ALS) [12]. Common SNPs in this gene family are also implicated in 18 complex human traits including hippocampal volume, type II diabetes, and aspects of Alzheimer's disease and lung cancer [13].

Recently thousands of edited sites were identified in bulk RNA sequencing of human tissues [14] as part of the GTEx consortium projects, including approximately 86,000 sites in the frontal cortex. The most well-characterised edited site is an A>I conversion at exonic nucleotide 2,135 of the glutamate receptor subunit transcript *GLUR2*, which produces a Q>R amino acid substitution. This site is edited in nearly 100% of human *GLUR2* transcripts, and limits the calcium permeability of the resulting ion channel, which is thought to dampen neuronal excitability [15]. Unlike this 'gold standard' *GLUR2* site, most editing sites are located in non-coding regions of the transcriptome, particularly within Alu repeats that form double-stranded RNA on which ADAR enzymes act. Editing of non-coding regions can lead to intronic retention, splice site variation and altered translation efficiency [16].

The RNA editing landscape has been described both in the healthy brain and in brains from neurological and neuropsychiatric patients. However this has been performed with bulk RNA sequencing methods, and whilst informative, little is known about the regional and cellular specificity of this process in single cells. Unlike gene expression analysis, RNA editing analysis requires greater sequencing depth and transcript coverage, which is more expensive than widely used 3' single-cell sequencing protocols. Picardi and colleagues previously profiled editing rates at known ADAR targets in 268 brain cells from 7 adult and 3 foetal brains [17,18] and found a markedly different distribution of editing site penetrance than is detected in bulk tissue sequencing. To better understand RNA editing dynamics in single brain cells, we analysed more than 3,000 neurons (16 subtypes from six cortical regions) from the left hemisphere of a single donor, previously reported by Lake and colleagues [19]. This dataset comprises the nuclear (largely pre-RNA) transcriptome, profiled using SMART-seq technology in order to maximize coverage across each transcript. We applied rigorous statistical methods to identify clinical and transcriptomic correlates of RNA editing in this dataset, and report site-specific differential editing across neuronal sub-types and cortical regions. This work represents the largest and most comprehensive analysis of RNA editing in single cells of any biological system to date.

MATERIALS and METHODS

RNA sequencing data and reference databases

Raw data from full-length (SMART-seq) sequencing of single brain cell nuclei was downloaded from dbGaP (study accession phs000834; project ID 20576) [19]. The neuronal sub-type annotations assigned by the study authors based on gene expression profiles (8 excitatory and 8 inhibitory neuronal sub-groups) were also downloaded. Single nucleus SMART-seq data for 116 neurons from the anterior temporal lobe of four adult donors was downloaded from the Gene Expression Omnibus (GSE67835) [18] using the NCBI SRA toolkit. Metadata including the cell type and cortical area of cell origin were accessed from the sequence read archive and/or relevant supplementary materials (STables 1 and 2). The RADAR database of RNA editing sites in Alu repeats, non-Alu repetitive regions, and non-repetitive regions was downloaded from http://lilab.stanford.edu/GokulR/database/ [20]. RADAR sites, edited sites in Brodmann area 9 reported by Tan et al [14]; and differentially edited sites from neuropsychiatric brain analyses [6,7], were transposed from Hg19 to Hg38 using liftOver software developed for the UCSC Genome Browser [21]. Genomic ranges for Alu and non-Alu repeats in Hg38 (RepeatMasker track) were downloaded from the UCSC Table Browser [22]; common human single nucleotide polymorphisms were downloaded from dbSNP (GRCh38p7; build 151) [23]; and clinically significant genomic variants were downloaded from NCBI (ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/; modified 29 August 2019) [24]. Tables of proteins evidenced to bind both DNA and RNA; small nucleolar RNAs and their host genes; genes edited in temporal lobe epilepsy patients; and transcripts with known amino acid recoding sites were retrieved from tables provided in the relevant literature [3,25–27].

RNA sequence processing and gene expression quantification

RNA sequencing reads were mapped to the human reference genome (GRCh38.91) using the STAR aligner [28] in two-pass mode with the splice junction database overhang maximized for the respective read lengths (47 for phs000834; 74 for GSE67835). Gene expression for the phs000834 dataset was quantified using featureCounts with fractional assignment of multi-mapping reads [29], filtered to remove samples with abundant mitochondrially-encoded RNAs (likely indicating apoptosis), low library complexity and low coverage. Count libraries were normalized for size and complexity, and clustered using scran [30] and scater [31]. For cluster analysis, nuclei were coloured according to the neuronal phenotype and cortical region designated by Lake and colleagues [19].

RNA editing detection and filtering

We called single nucleotide polymorphisms (SNPs) in the nuclear transcriptomes using the GATK best-practices pipeline [32] for RNA-seq data. This involved read de-duplication, splitting, base quality score recalibration, and variant calling with HaplotypeCaller. No call confidence filter was applied at this step in order to retain all potential variants. Variant call format files were converted to GDS format using the R SeqArray package [33], and combined with common genomic SNPs, and RADAR sites. For phs000834, A>G and T>C sites covered by at least 5 reads, with a minor allele count of at least 2, and detected in at least 10 high-quality (i.e., non-apoptotic, high transcriptional complexity) libraries, were retained. As SMART-seq data is unstranded, categorically determining the originally edited strand is not possible. To minimize false positive sites, we therefore imposed further filtering as follows. Common genomic SNPs were discarded, and sites were only retained if: a) they were present in previously published RNA editing databases, or, b) located within non-overlapping regions of an ensemble feature, and on the cognate strand (i.e. A>G sites within a feature on the 5' strand; T>C on the 3' strand). The bedtools intersect module was used to locate previously unreported sites within genomic repeats (RepeatMasker); and to identify the genic feature (e.g. start codon, exon, intron, 3' UTR etc) occupied by sites within protein-coding genes [34]. Sites with insufficient coverage were distinguished from those covered by at least five reads but with little to no alternate allele present, using the samtools depth module [35]. The Variant Effect Predictor [36] was used to predict the molecular consequences of RNA editing sites. For GSE67835 the same site detection and filtering procedure was followed except no threshold for prevalence across nuclei was applied, given the smaller number of neurons under consideration.

Statistical analysis

Editing sites that survived filtering were cross-referenced with sites reported in bulk RNA sequencing from the frontal cortex (Brodmann area 9) [14], and those identified in single neuronal nuclei prepared from four donors (GSE67835) [18]. Data transformation and analysis was performed in R [37], using the dplyr [38] and tidyr [39] packages. doParallel [40] and multidplyr [41] were used for parallel processing. Data were visualized using ggplot2 [42], UpSetR [43] and circlize packages [44].
Neuron-level 'global' editing proportions were quantified as the number of edited sites, relative to the total number of transcribed candidate editing sites in each cell cell. Sites supported by at least five reads were considered to be transcribed. Edited sites were subsequently grouped by transcriptomic context (Alu repeats, repetitive non-Alu sequence, and non-repetitive sequence) to determine the contribution of each to global editing. Differences in proportionate RNA editing across neuronal groups and cortical regions were tested using linear models and the R broom package [45].

Relationships between the log-transformed neuronal transcript abundance (transcripts per million) and neuronal editing proportion were assessed with linear models, controlling for gross neuronal phenotype (excitatory or inhibitory) unless otherwise stated. Gene set Gene Ontology enrichment analysis was performed using the limma goana module [46], after exclusion of genes in which edited sites were detected, and those with effect size estimates less than 0.01. For tests of interaction between neuronal phenotype, gene expression and editing, results with fewer than 30 cells per neuronal group were discarded.

Site-specific differential editing was quantified across gross neuronal phenotypes within 6 cortical regions (i.e., 12 experimental groups) using chi square tests of independence, employing the permutation testing procedure (999 iterations) when expected cell counts fell below 5. All results were corrected for multiple testing using the Benjamini-Hochberg method (FDR < 0.05). Post-

134 hoc pair-wise testing of significant differences was performed using the pairwiseNominalIndependence test from the rcompanion
135 package [47], with FDR < 0.05.
136
137 **RESULTS**
138
139 *Thousands of novel putative RNA editing sites revealed at single-cell resolution*
140
141 Lake and colleagues previously reported the transcriptional features of over 3000 high-quality neuronal nuclei isolated from a
142 single donor, sequenced using the SMART-seq full-length transcript sequencing platform [19]phs000834; [19]. Our independent
143 read processing and unsupervised clustering of these data separated nuclei according to reported neuronal sub-types (SFigure
144 1). We discarded 72 of the 3,127 neuronal nuclei due to either high abundance of mitochondrial reads (n=42), or low library
145 complexity (30) (read mapping statistics are provided in SFigure 2). We therefore quantified RNA editing signals in a total of 3,055
146 neuronal nuclear transcriptomes. Excitatory neurons from BA41 were the most abundantly represented cell type in this data set
147 (606 cells; 29.7% of total), and the mean proportion of inhibitory neurons across each cortical region was 28.5% (Figure 3b).
148
149 Filtering on site coverage, prevalence, genomic context, and previous evidence, yielded 884,995 RNA editing records of 40,861
150 unique sites within 5,880 genes. An average of 1,808 candidate editing sites were transcribed per cell, of which 287 were edited
151 (Figure 1a). The median number of cells in which an edited site was detected was 15 (Figure 1b). Approximately 83% of edited
152 sites were located within non-overlapping regions of protein-coding genes. The majority of these (53%) were present in a curated
153 database of human RNA editing sites [20], and are located in intronic Alu repeats (Figure 1c & d). A further 3,074 sites were
154 exonic, followed by sites in three-prime untranslated regions (2,080), five-prime UTRs (77) and stop codons (4), in broad
155 agreement with previous genome-wide characterization of RNA editing [8]. The 15,784 previously unreported sites tended to be
156 edited in fewer cells than documented sites (SFigure 3), and comprized 8,845 sites within Alu repeats; 565 in repetitive, non-Alu
157 regions, and 6,374 in non-repetitive regions (Figure 1d). Sites in coding regions most frequently occupied codon position 2,
158 downstream of cytosine (SFigure 4).

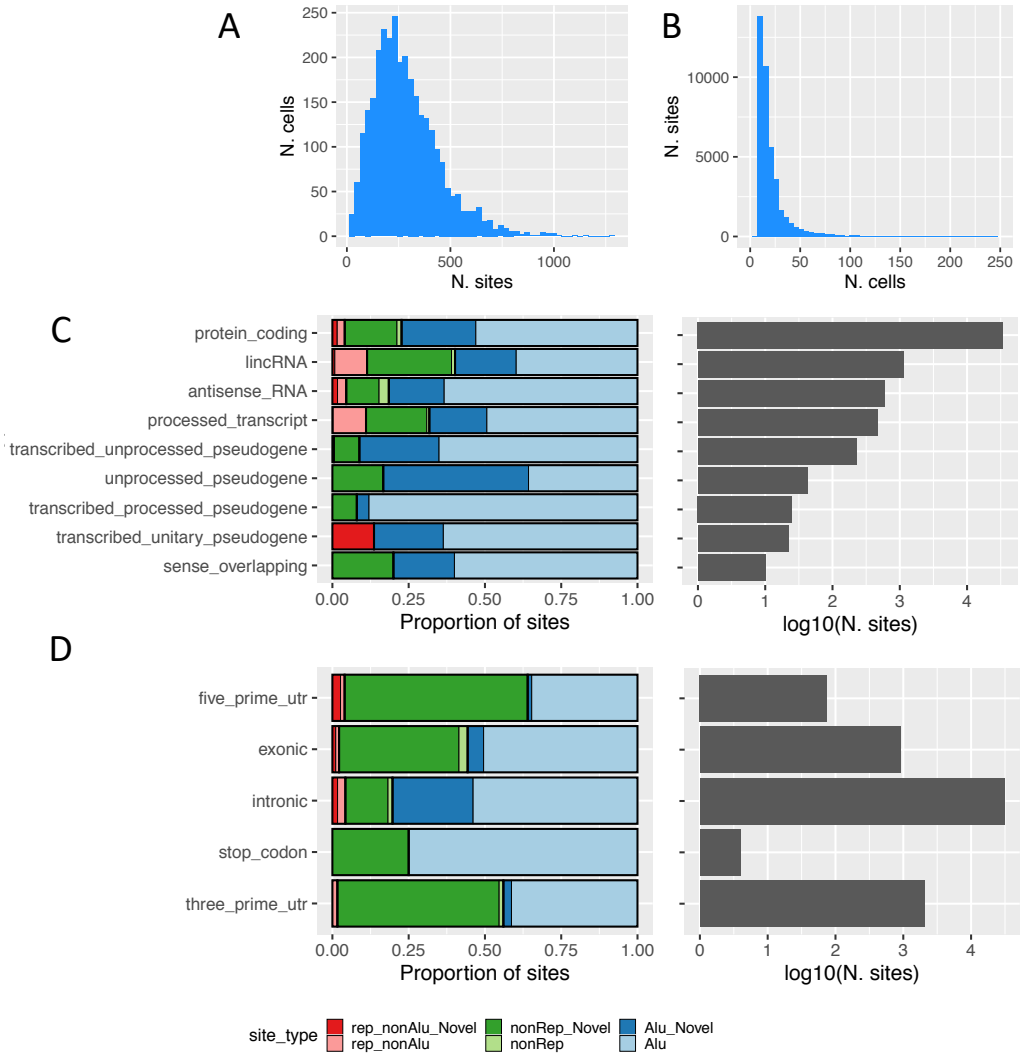
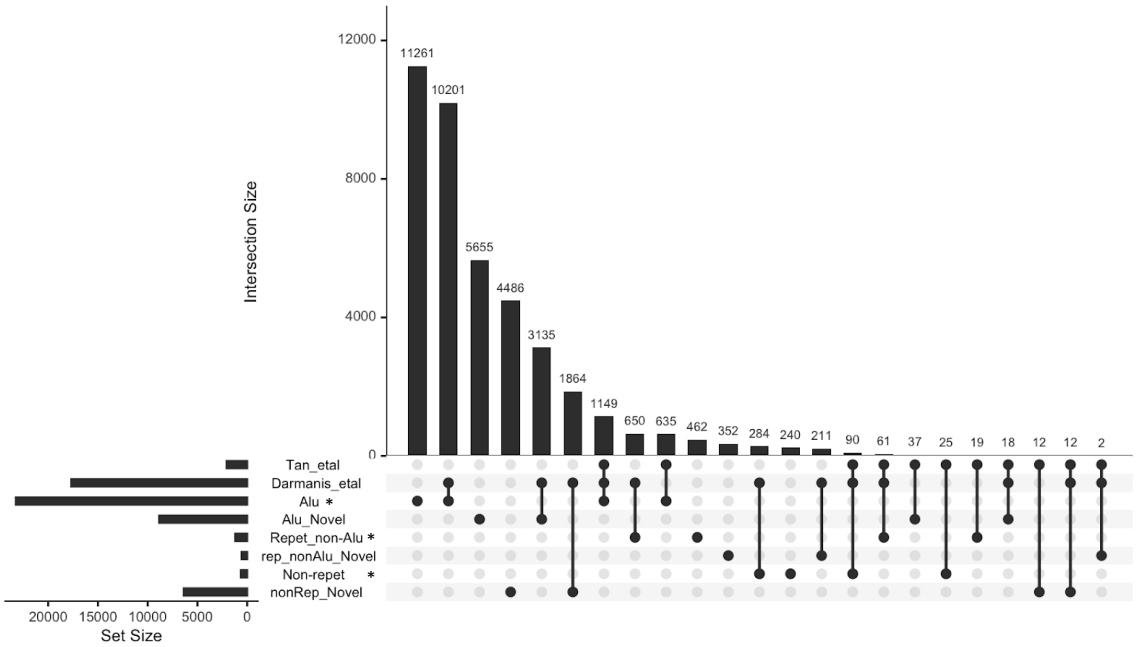


Figure 1. Distribution and genomic context of edited sites across 3,055 neuronal transcriptomes. A) Distribution of the number of sites detected per neuron. B) Distribution of the number of neurons in which each site is detected. C) Genomic context of edited sites by gene biotype and sequence type, as defined in the RADAR database (left panel). Novel sites denote those not present in RADAR. Total number of sites in each category (right panel). Genomic features with fewer than 5 associated editing sites are omitted for brevity. D) Genic features in which sites in protein-coding genes are located (left panel). This figure represents only sites in non-overlapping gene regions (83% of total). Total number of sites in each category (right panel). A summary of all detected sites is provided in Supplementary Table 3.

We investigated agreement between sites identified in these neurons with RNA editing sites identified in bulk RNA sequencing of frontal cortex (BA9) from 70 healthy donors [14], and in single-nucleus SMART-seq libraries corresponding to 111 neuronal nuclei from four unrelated donors [18] (STable 2).

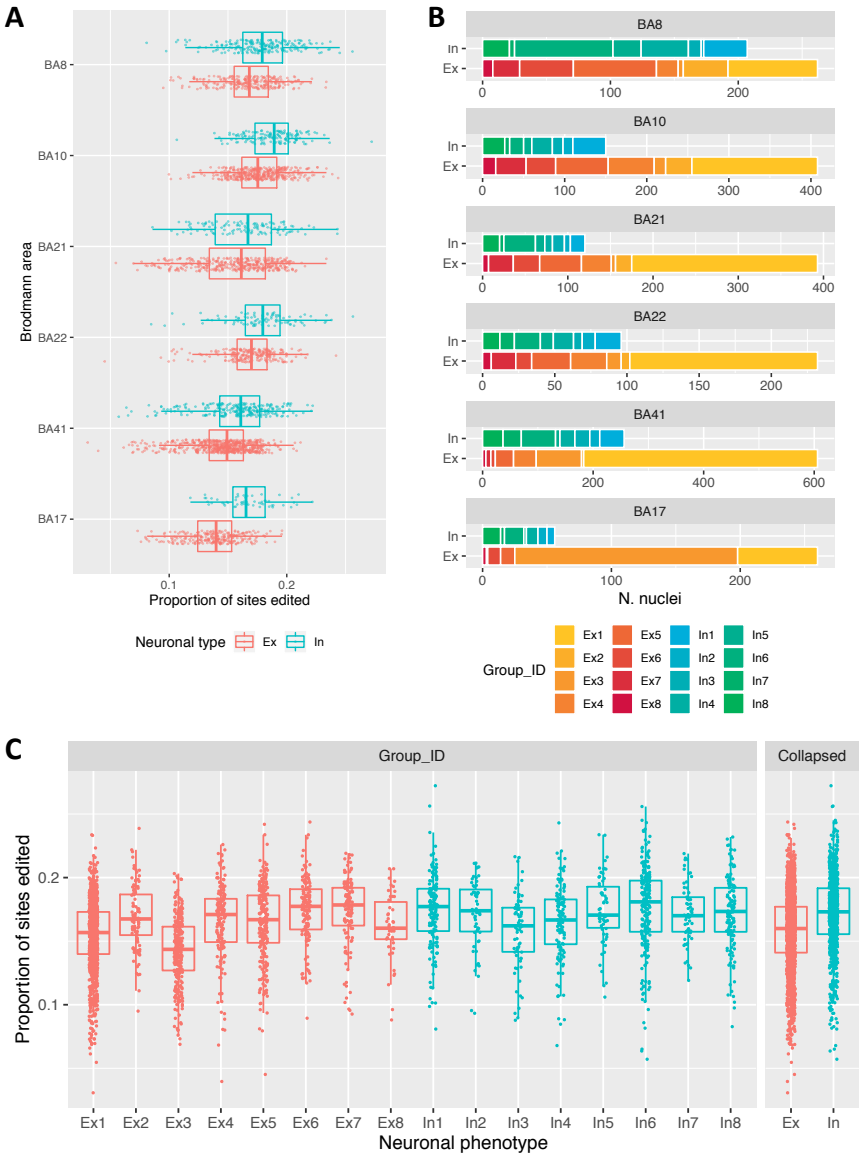
Some 17,677 sites (43% of total) identified in the sole donor (phs000834) were also detected in the neuronal nuclei of at least one donor in GSE67835 [18], including 1,876 (30%) of 6,374 novel non-repetitive sites. However only 2,060 sites agreed between the sole donor and those reported from bulk RNA sequencing of Brodmann area 9 [14] (Figure 2). We further intersected our

176 data with differentially edited sites reported after bulk brain RNA sequencing of neuropsychiatric patient cohorts, and could
177 provide single-cell level context for 1,871 (10% of total) differentially edited sites in the frontal cortex of schizophrenic patients
178 [7]; and 998 (16.5% of total) differentially edited sites in autism spectrum disorder patients [6] (STable 3).
179 Taken together, the majority of sites detected in phs000834 were also present in public databases of known RNA editing sites,
180 and/or recent single cell and bulk brain RNA sequencing editing data sets. This demonstrates that RNA editing is biologically
181 conserved and can be reproducibly detected across different individuals with the same scRNA-seq platform. Signals detectable
182 in single-cell data however, are not as reproducibly detected in bulk RNAseq, suggesting highly cell-specific patterns of RNA
183 editing which are subsumed in the latter data type.
184



185
186 **Figure 2. Intersection of sites detected in the single donor phs000834 with sites in related data-sets.** Asterisks
187 denote sites present in the RADAR database (Ramaswami and Li, 2013), which are mutually exclusive of each other,
188 as are novel sites. Note that although 'novel' denotes absence from the RADAR database, many such sites are also
189 detected in independent studies of bulk brain (Tan et al. 2017) (top row) and single neuron transcriptomes (Darmanis
190 et al. 2015) (second top row).
191
192
193 *Small nucleolar RNA expression is a marker of nuclear RNA editing*
194
195 We examined the alternate (i.e., edited) allele proportion for each site and found a strong bias towards 1.0 (**SFigure 5**), indicating
196 total replacement of the genomic allele with the edited site. This was previously observed in analysis of editing in small numbers
197 of brain cells from 7 donors (GSE67835) [17,18]. Therefore in order to quantify editing in each neuronal nucleus, we binarized
198 the editing signal at each transcribed site to 'edited' or 'non-edited', and calculated the proportion of edited sites. We hereafter
199 refer to this proportion as the cell-wise 'global editing index' (GEI).
200
201 Overall, inhibitory neurons exhibited significantly greater mean GEI than excitatory neurons, which was partly attributable to
202 greater editing in In6 and lower editing in Ex3 sub-groups (Figure 3c). Differences in mean GEI between cortical regions was

203 related to different proportions of neuronal sub-types (Figure 3a & b). Specifically, the visual cortex (BA17) was enriched for
204 neurons of sub-group Ex3, and showed both the lowest proportion of inhibitory neurons (18%), and the lowest mean GEI (see
205 **STable 4** for linear modelling results). Similarly the superior temporal cortex (BA41) was enriched for group Ex1 neurons, and
206 showed a lower mean GEI than the frontal cortex (BA8 and BA10) in which inhibitory neurons (In1, In6 and In8) were more
207 abundant.
208



209
210 **Figure 3. Editing differences between cortical regions and neuronal sub-types.** A) Distribution of global editing
211 index (GEI; i.e., edited sites relative to all transcribed candidate sites for each cell) across six cortical Brodmann areas
212 (BA), for inhibitory (teal) and excitatory (red) neurons. All groups except for Ex neurons in BA10 and In neurons in
213 BA22 are show significantly different GEI compared to In neurons of BA8. B) Abundance of neuronal sub-groups
214 between cortical regions, as defined by Lake and colleagues (Lake et al. 2016). Note different x axis scales. C) Global
215 editing index in neuronal sub-groups, and when collapsed into gross neuronal phenotypes. All sub-groups are
216 significantly different to Ex1 except for Ex8 and In3. In: inhibitory neuron; Ex: excitatory neuron.
217
218

Previous studies identified novel modulators of RNA editing by correlating transcript abundance and global editing activity across multiple tissue isolates [14]. We used a similar approach at the single-cell level to identify strong correlates of the cell-wise GEI. This analysis was controlled for gross neuronal type (inhibitory or excitatory), and excluded transcripts with absolute effect sizes less than 0.01, and directly edited transcripts — with the following exceptions. Although 32 edited sites were detected in transcripts encoding the *ADAR* gene family (28 in *ADAR3*), we directly tested the relationship between expression of these genes and cell-wise editing proportion, controlling for neuronal phenotype. *ADAR3* expression correlated negatively with editing ($p = 0.003$), and surprisingly *ADAR1* and *ADAR2* also showed modest nominal negative effects on GEI ($p = 0.03$ and 0.02 respectively; not significant after FDR correction). We further examined these results by splitting the GEI into three components, derived from Alu sites, repetitive non-Alu sites, and non-repetitive sites in the manner of Ramaswami and colleagues [20]. We tested the correlations between each component and *ADAR* expression for both inhibitory, and excitatory neurons. After correcting for multiple testing, positive correlations remained between *ADAR1* expression and editing at non-repetitive sites in excitatory neurons, and at repetitive non-Alu sites in both neuronal types (Figure 4a & b). Interestingly *ADAR2* expression was positively correlated with editing at the same site types, but only in inhibitory neurons. Conversely *ADAR3* expression was negatively correlated with editing at Alu sites in excitatory neurons, and at non-repetitive sites in inhibitory neurons (Figure 4a).

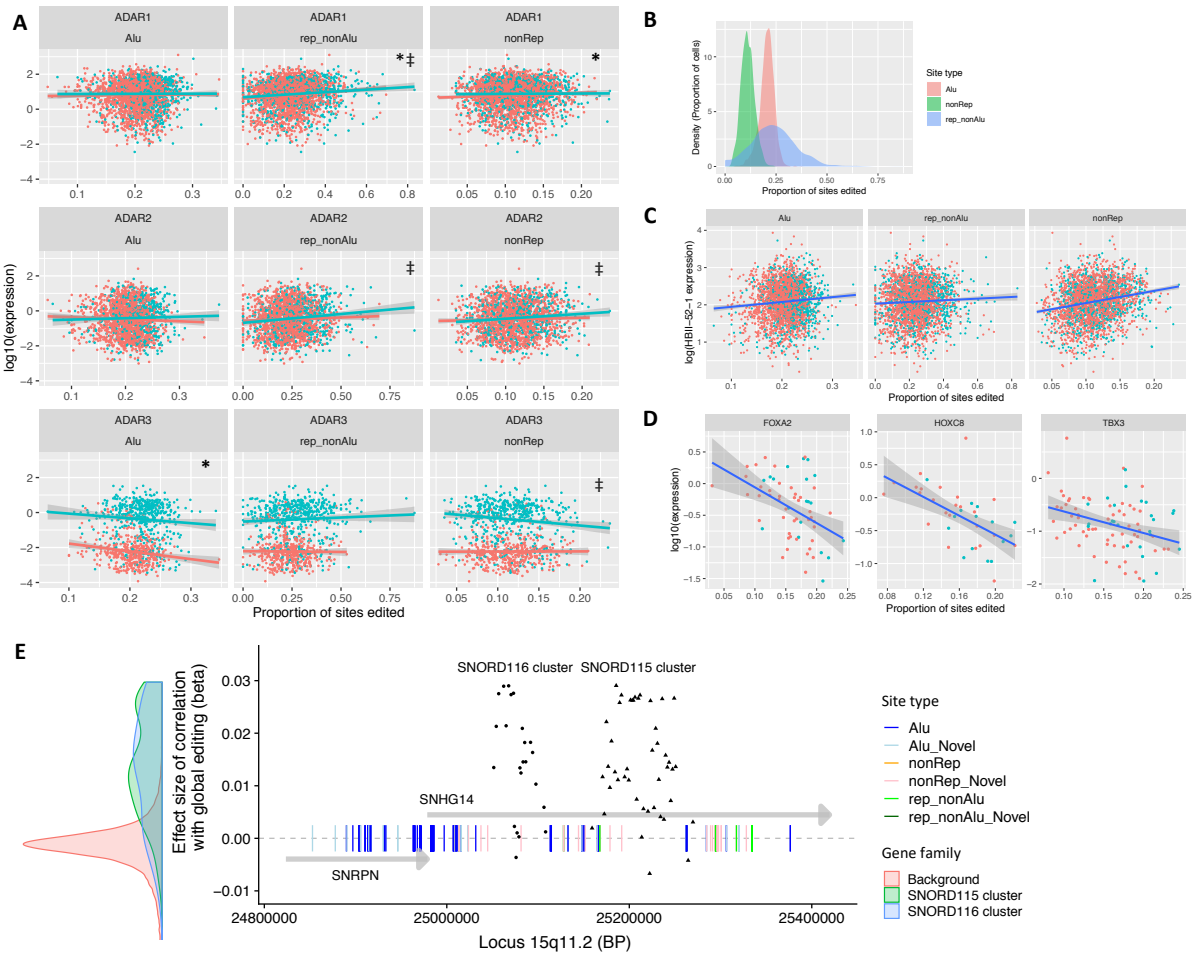


Figure 4. Gene expression correlates of global RNA editing. A) Expression of ADAR family enzymes (y axis) relative to global editing proportions in excitatory (red) and inhibitory (teal) neurons (x axis). Editing proportion, calculated as the ratio of edited to transcribed sites in each nucleus, was divided according to the genomic context of the edited sites – within Alu, repetitive-nonAlu, and non-repetitive sequence. Significance markers represent corrected $p < 0.05$. * significant correlation between editing proportion and excitatory neurons; ‡: significant correlation between editing proportion and inhibitory neurons. B) Distribution of editing proportions per neuron, according to genomic context. C) Editing proportion per genomic context (x axis), relative to expression of a representative small nucleolar RNA from the SNORD115 cluster, HBII-52 (ENSG00000201831; y axis). D) Global editing proportions (x axis) relative to expression of representative FOX, HOX and T-BOX gene family members. E) Expression of a cluster of small nucleolar RNAs in locus 15q11.2 correlates strongly with global RNA editing. Genomic location is displayed in base pair units on the x axis; effect size of correlation with editing is displayed on the y axis for SNRPN, SNHG14 (grey arrows) and SNORD cluster genes (black points). The distribution of correlations between the SNORD clusters and editing is displayed at left relative to all other genes (background), and scaled to the y axis. Edited sites are represented by vertical lines, and coloured according to genomic context.

After controlling for neuronal phenotype, negative correlations between transcription and GEI were substantially more frequent than positive correlations (844 and 138 respectively). A further 222 transcripts showed neuronal type-dependent associations with editing (**STable 5**). Among genes positively correlated with GEI, those involved in RNA processing were the most enriched gene ontology (GO) category, consisting of 54 small nucleolar RNA genes (C/D box 115-116 on chromosome 15; and C/D box 3A on chromosome 17). Another 50 non-coding transcripts including pseudogenes, antisense and ribosomal RNA, and long interspersed ncRNA, were positively correlated with GEI (36% of all positively correlated genes). Protein-coding genes in this group were involved in nucleic acid binding (*CHD5*; and the transcription factor *MESP2*), solute transport (*SLC25A20*), the ubiquitin system (*ZNRF3*), and glycosyltransferase activity (*GLT6D1*) (STable 5). To further investigate the strong correlation between snoRNA transcription and editing, we examined the correlation between each GEI component (detailed above) and transcription of *HBII-52*, a representative C/D box snoRNA in the SNORD115 cluster. Editing at non-repetitive sites was the strongest correlate of transcription with *HBII-52* (Figure 4c). Importantly, this result was robust to removal of a large cluster of edited sites in the small nucleolar RNA host gene 14 (enclosing the SNORD115 and -116 clusters) from the editing proportion calculations (Figure 4e; SFigure 6).

Transcripts negatively associated with editing were enriched for GO terms such as 'regulation of biosynthetic process' (forkhead box, T-box, GATA and homeobox DNA binding proteins); 'signaling receptor activity' (chemokine receptors; the GABA receptor rho1 sub-unit, several olfactory receptor families, and TNF and toll-like receptors); and neuronal differentiation (LIM homeobox; POU homeobox; synaptosome-associated protein 25 and WNT2) (STable 6) (Figure 4d). Together these results represent multiple putative novel modulators of RNA editing, which await further exploration in the context of neuronal differentiation and function.

Finally, when transcriptional correlates of editing were allowed to include directly edited transcripts and those with small effect sizes, we reproduced 35 of 147 negative transcriptional correlates of editing reported by Tan and colleagues in bulk RNA sequencing of the frontal cortex [14]; and 8 of 144 positive correlations. When RNA binding proteins that modulate RNA editing in cell cultures were considered [48], 17 of 107 reported inhibitors also showed negative correlations with editing in our data. Lastly, further investigation of the numerous DNA-binding proteins associated with editing, revealed 28 with both DNA- and RNA-binding capacity reviewed previously [25] (STable 5).

281 *Predicted functional effects and clinical associations of edited sites in single neurons*
282
283 Of 40,861 detected editing sites, functional effects could be predicted for 5,907 sites (14.4% of total; excluding up/down-stream
284 gene variants, intergenic and intronic variants), with 327 sites linked to amino acid substitutions. We found edited sites in ten
285 well-established targets of ADAR enzymes that produce modified peptides upon editing, of which four (in *NEIL1*, *GRIK2*, *CYFIP2*,
286 and the 'gold standard' site in *GLUR2*) had missense mutations which were identical to those previously reported [26] (STable 7).
287 For 295 missense sites in non-overlapping regions of 229 genes, the most enriched gene ontology terms included 'adenyl
288 ribonucleotide binding' (including DEAD box helicases, phenylalanyl tRNA synthetase 2, HSP90 alpha AA1 and AB1, and several
289 calcium transporting ATPases), 'neurotransmitter secretion' (piccolo presynaptic matrix protein, Ras homologue *RHOT1*, syntaxin
290 1B and related binding proteins), and 'brain development' (NMDA glutamate receptor 1; C5orf42, E3 ubiquitin ligase *HERC1*,
291 myosin, neurofibromin and GTPase family members).
292
293 Interestingly, six previously unreported sites coincide with genomic nucleotides whose mutations are associated with specific
294 clinical syndromes when queried in the ClinVar database (STable 8). Of particular interest from a neurological perspective are five
295 SNPs that are predicted to disrupt editing in the *DDOST* transcript (encoding a protein glycosyltransferase subunit), associated
296 with Parkinson's disease; in 24-dehydrocholesterol reductase, associated with desmosterolosis; in phenylalynyl-tRNA synthetase
297 2, associated with oxidative phosphorylation deficiency; in myosin VI, associated with hearing loss; and in *SIGMAR1*, associated
298 with ALS [49]. The sixth clinically relevant variant result is an amino acid substitution implicated in tachycardia (MIM 180902).
299 More information is now required for these associations, which were mostly derived through unpublished clinical sequencing
300 results (Illumina clinical services).
301
302 *Differential site editing across cortical regions and neuronal sub-groups*
303
304 Given the differing proportions of excitatory and inhibitory neurons that were sampled across the six cortical regions in the
305 original study (Figure 3b), we tested site-specific differential editing between cortical regions, and the neuronal subtypes therein.
306 Specifically, we discretized 40,861 sites as 'edited', 'non-edited' or 'not-transcribed' in each nucleus. We then performed chi-
307 squared tests of independence on the first two categories across nuclei at each site, grouped by cortical region and gross neuronal
308 phenotype.
309 Some 8,300 sites were differentially edited (DEd) between neuronal types across cortical regions, of which 117 sites within 53
310 genes were DEd in 353 pair-wise post-hoc tests (FDR=0.05). Differential editing between the excitatory and inhibitory neurons of
311 the frontal and temporal cortices was the most common finding (83 DEd results; 23.2% of total). Differential editing between
312 excitatory neurons of the frontal and temporal cortex was also common (14.6%) followed by differences between neuronal sub-
313 types across the three temporal cortical regions (11.2 %) (Figure 5a).
314

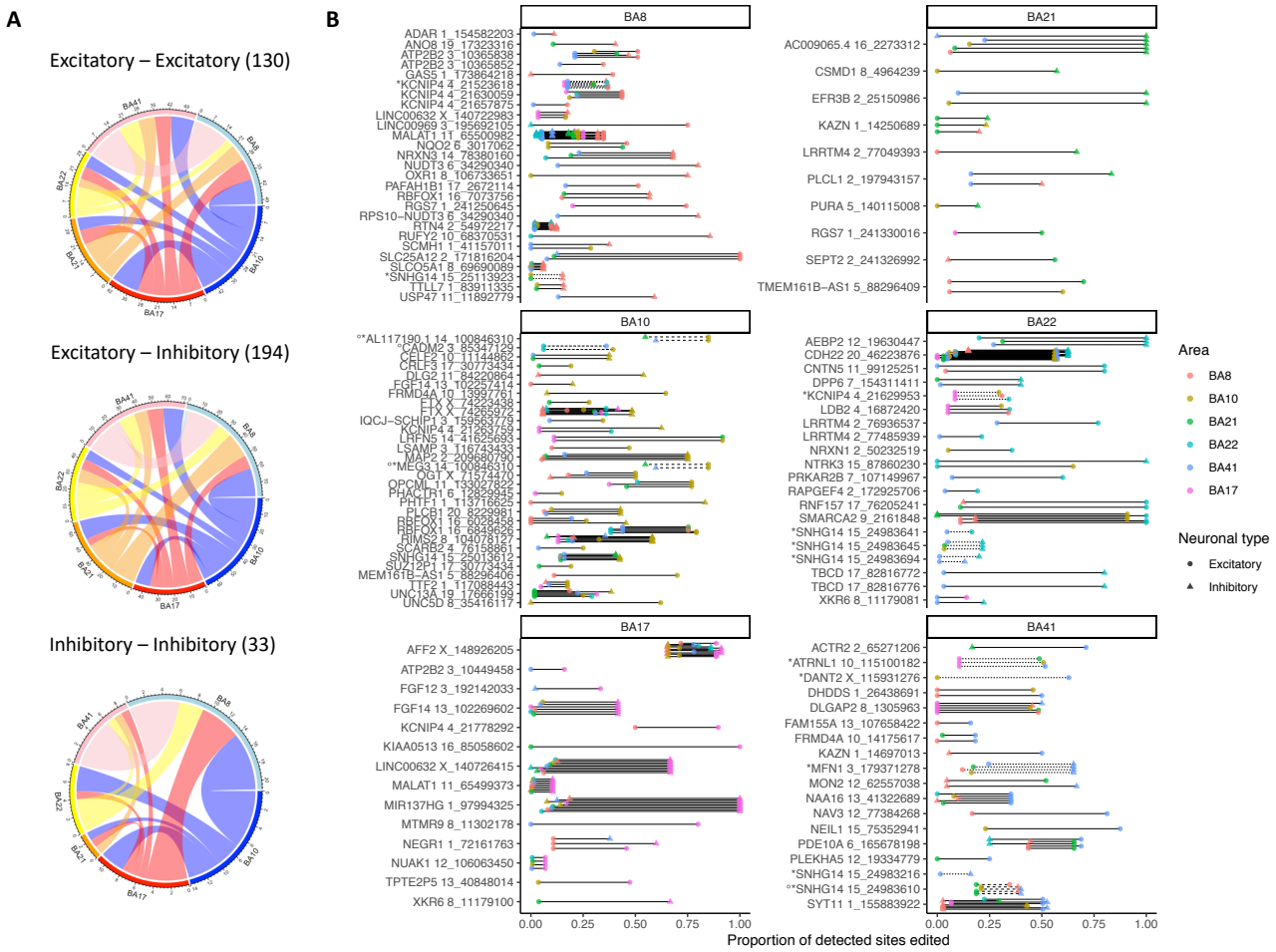


Figure 5. Differential editing at individual sites between neuronal cell types and cortical regions. A) Differentially edited sites between excitatory neurons (top), excitatory and inhibitory neurons (middle) and inhibitory neurons (bottom panel) between cortical regions. The total number of differentially editing results is displayed in parentheses. B) Proportionate editing for differentially edited sites in panel A, after in pair-wise comparisons of cortical regions and constituent neurons (x axis). Gene symbol, chromosome and location (Hg38 reference) are displayed on the y axis. Brodmann area is represented in the point colour, and gross neuronal phenotype is encoded in the point shape. Sites are grouped according to the cortical region in which the greatest proportionate editing is observed. Sites differentially edited in the frontal and/or temporal cortex of schizophrenic and autism spectrum disorder patients, are respectively denoted with asterisks and circles in the y-axis labels, and by different line types in the plotting space.

Of note among sites within non-coding genes were six of seven sites in Alu repeats of the small RNA host gene 14 (chromosome 15) that were more edited in inhibitory neurons in multiple frontal and temporal regions, relative to co-located excitatory neurons. Two sites in the long non-coding RNA 632 (X chromosome) were relatively more edited in inhibitory neurons of the frontal and temporal cortices respectively; whereas a site in *LINC969* (chromosome 3) was preferentially edited in excitatory neurons in the frontal cortex. Sites within transcripts related to neuronal differentiation and maturation were consistently differentially edited in our results (**STable 9**). For example, *RBFOX1* (involved in splicing and maturation) exhibited two DEd sites, enriched in excitatory neurons in the temporal cortex, and inhibitory neurons of the frontal cortex, respectively. The neural chromatin remodelling factor *SMARCA2* was consistently edited in excitatory neurons of BA10 (frontal lobe) and BA21 (temporal

337 lobe) relative to inhibitory neurons in those lobes. Two sites within a synaptic membrane protein transcript, *LRRTM4*, were DEd
338 only amongst excitatory neurons within the temporal lobe; whereas synaptotagmin 11, which also localizes to the synapse, was
339 preferentially edited in excitatory and inhibitory neurons of the temporal cortex relative to the frontal and visual cortex. In terms
340 of ion channels, the potassium channel interacting protein 4 was differentially edited at six sites, with greater proportional editing
341 in the frontal cortical regions (BA8 and BA10; both neuronal types) than BA17 and BA41. Finally, second-messenger associated
342 enzyme transcripts phospholipase C-B1 and -L1 were preferentially edited in inhibitory neurons of the frontal and temporal
343 cortices respectively, whereas phosphodiesterase 10A was more edited in the excitatory neurons of the temporal cortex than
344 other regions. A complete summary of these findings is presented in Figure 5b, with editing statistics in STable 10.
345
346 *Differentially edited sites in neuropsychiatric patients show cell type-specific editing*
347
348 We next compared the DEd sites in our single nuclei data to sites that are reported to be differentially edited in cortical samples
349 from neuropsychiatric patients compared to controls (Figure 5b). Two sites in *KNCIP4*, and sites in the small RNA host gene
350 *SNHG14*, the long non-coding RNA *MEG3*, were DEd in the frontal cortex of schizophrenic patients (SCZ) relative to controls, and
351 in our data, exhibited relatively more editing in frontal cortical neurons than those in the temporal and visual cortices. The cell
352 adhesion molecule CADM2 was DEd in the frontal cortex of patients with autism spectrum disorder (ASD) and in excitatory frontal
353 cortical neurons in our data.
354
355 Five additional sites in *SNHG14* were DEd in SCZ, and preferentially edited in inhibitory neurons of the temporal cortex relative
356 to excitatory neurons in temporal and frontal regions. In particular, the *SNHG14* site 15:24983610 was more highly edited in
357 inhibitory neurons of the frontal and temporal cortex in our results, and was DEd in two different frontal cortical regions from
358 SCZ patients, a validation SCZ patient cohort, and the frontal cortex of ASD patients. Finally the attractin like protein 1 transcript
359 and ncRNA *DANT2* were preferentially edited in excitatory neurons of the temporal cortex compared to visual and frontal cortical
360 regions respectively in our data, and DEd in the frontal cortex of SCZ patients; while mitofusin 1 was more edited in the same
361 patient brain regions, and in inhibitory neurons of the temporal cortex in our data (Figure 5b). Relaxing the pair-wise DEd
362 significance threshold to FDR=0.1 yielded another 15 sites DEd in both neuropsychiatric cohorts, and different regions and cell
363 types in our data (STable 10).
364
365 **DISCUSSION**
366
367 RNA editing is ubiquitous in the human central nervous system, and aberrant editing is implicated in numerous
368 neurological/neuropsychiatric diseases, cancer and immune disorders. We undertook the first high-resolution investigation of
369 RNA editing in full-length RNA from single neuronal nuclei across 6 cortical regions, and found extensive agreement with
370 previously reported edited sites in healthy and disease-affected cohorts demonstrating that RNA editing analysis is both: (i)
371 feasible in scRNAseq, and (ii) able to detect disease-relevant sites, but with the added benefit of allowing investigation of cell
372 type specificity. The majority of edited sites were within intronic or 3' UTR Alu repeats, and those within coding regions tended
373 to occur immediately downstream of cytosine. These characteristics are in broad agreement with those reported previously for
374 edited sites in the human transcriptome [8].
375
376 This work was made possible by publicly available sequencing data and curated databases. The novel sites identified here require
377 further validation and functional studies, and will contribute to the development of much-needed cell-type specific editing
378 reference databases. The main limitations of this study lie in the lack of RNA stranding information in the SMART-seq libraries,

379 and the lack of reference genomes for brain donors— both of which hamper our ability to disambiguate editing sites from
380 genomic SNPs. To address these concerns we took several measures including: i) removing common genomic SNPs from the
381 results, ii) removing sites in transcripts derived from overlapping gene regions, iii) retaining only sites within genes on the cognate
382 strand, and vi) cross-referencing our data with sites detected in single neuronal nuclear transcriptomes from four additional
383 unrelated individuals, sequenced with the same technology [18]. Further, because we retained only sites detected in at least ten
384 nuclei, PCR artifacts should have minimal influence on our results.
385
386 Excitingly, thousands of putative novel editing sites were detectable in the present data, thanks to the extensive coverage of the
387 transcriptome at single cell resolution. That a substantial proportion (30-38%) of 15,784 newly reported sites in this study were
388 also detectable in independent SMART sequencing of small numbers of neurons from unrelated individuals, suggests that this
389 data type contains novel and reproducible insights which are not detectable in bulk RNA sequencing. Indeed, the cellular
390 heterogeneity of RNA editing is a striking result in the present study. Whereas previous studies of bulk RNA portray editing as a
391 generalized, low-penetrance phenomenon [14], our results suggest that this is, at least in part, an artefact of highly penetrant
392 but cell-specific processes that are occurring within bulk tissue, but masked due to heterogeneity. In addition, the present data
393 is enriched for nuclear pre-RNA, and may predominantly reflect ADAR2-dependent editing. ADAR2 is spatially restricted to the
394 nucleus, and predominantly expressed in neurons. These differences may account for the relatively poor agreement between
395 sites reported in bulk RNA-sequencing from cortical tissue [14], and those reported here. Conversely, highly prevalent but low-
396 penetrance editing, and sites edited in cytosolic RNA, will be under-represented in the current data.
397
398 *Transcriptional correlates of editing*
399
400 Given that individual edited sites were generally restricted to a small number of cells in our data, we performed analyses using
401 the proportion of edited sites as a ratio of all transcribed candidate sites per cell, to assess impact on gene transcription, revealing
402 new relationships. ADAR3 (also known as ADAR2B) expression was negatively correlated with editing as expected, however the
403 effect of ADAR1 and 2 expression on editing appeared confounded by the genomic context of the edited sites. When the editing
404 signal was divided according to genomic context, a positive relationship between ADAR1 expression and non-Alu site editing was
405 apparent. Interestingly ADAR2 expression correlated with non-Alu site editing in inhibitory neurons specifically, whereas the
406 major correlate of Alu site editing was expression of the inhibitory enzyme ADAR3. The latter enzyme is a marker of inhibitory
407 neurons [19,50], which showed greater proportional editing in our results. However despite its low expression in excitatory
408 neurons, our results indicate ADAR3 may yet have a role in repressing Alu editing in these cells. Variation in ADAR expression, by
409 contrast, may have a greater influence on editing in inhibitory neurons.
410
411 The modest correlations between ADAR family expression and global editing rates is not entirely unexpected. A study with more
412 than 8,000 bulk RNAseq samples from 552 individuals found that at most, 25% of the variation in RNA editing was attributable to
413 ADAR expression [14]. Indeed, numerous other proteins are reported to interact with ADAR members and influence editing
414 [14,48], many of which were replicated in our study.
415
416 Unlike previous studies, we found several associations between non-coding gene products and editing, most notably a major
417 positive correlation between small nucleolar RNA (snoRNA) abundance and global RNA editing. These transcripts are excised from
418 introns of coding and non-coding 'host genes', and form a scaffold on which ribonucleoprotein complexes form to direct post-
419 transcriptional modification of target RNAs [51]. Interestingly, ADAR2 is localized exclusively in the neuronal nucleolus and targets
420 non-repetitive sites [11,26]. Therefore the strong correlation we detected between snoRNA expression and editing at targets of

ADAR2 in particular (Figure 4c), suggests that enhanced snoRNA expression is a marker of ADAR2-mediated editing in the nucleolus (discussed below).

Consistent with previous studies, we found negative correlations between editing and poly-adenosine binding proteins and HNRNP RNA binding proteins. The negative association with multiple DNA binding protein families was unexpected, and cannot be explained by transcriptional repressor activity, as our measure of editing was adjusted for transcription. Dual DNA and RNA binding capacity is evidenced for more than 100 nucleic acid binding proteins [25] , including members of T-BOX, HNRNP, chromo- and homeobox family members which, together with several metabolic enzymes, were negatively associated with editing in our data. These findings may therefore derive from DNA-binding proteins sequestering nascent RNA, or competing for RNA access with ADARs. Further protein-RNA interaction experiments in neuronal stem cells are now required to validate these novel insights.

Site-specific editing

Our study was able to quantify editing sites across cortical regions and neuronal sub-types with high resolution, as well as investigate the transcriptomic underpinnings of neurological and psychiatric illness in the context of brain region and cell-specific editing. In fact, a major advantage of the present data set is the breadth of cortical regions under consideration. Recent publication of larger numbers of neurons from limited cortical regions will certainly assist in discovering and validating RNA editing sites in neuronal sub-groups, however the present data remains the most spatially comprehensive that is available to our knowledge.

We found that a number of transcripts encoding putative novel missense sites are involved in adenosine binding. Editing-based feedback loops in editing-related enzymes are reported in model systems [52], and may be more extensive than previously thought. Beyond adenosine binding proteins, the presence of putative recoding sites in synaptic transmission-related proteins, as well as solute carrier family members and neurotransmitter receptors, suggests that editing-based peptide modification modulates both the sensitivity of neurons to stimuli, and the responsivity at the synapse.

The present work implicates RNA editing in the pathogenesis of five rare diseases. A genomic A>G mutation at chr9:34635601 in the 3' untranslated region of SIGMAR1 causes ALS [49]. The mutation is predicted to create a splicing site, resulting in constitutive inclusion of additional exonic sequence, and presumably, a longer SIGMAR1 product. Our results suggest that RNA editing at this site may be a mechanism for alternate splicing, and that both isoforms are required for neuronal health. As an aside, the genomic adenosine chr1:237503413 is edited in our data, whereas its guanosine variant is implicated in catecholaminergic tachycardia. Editing at coding sites in the arteries is vastly higher than in other tissues [14] and could inform the cause of this tachycardia. Unfortunately, little functional data is currently available for the remaining clinical conditions in which our results suggest RNA editing may be involved. Nevertheless we note that all six candidate disease-related editing sites are also detected in single neurons of at least one other individual [18]. We further envisage that the detailed editing catalogue reported here will inform future investigations into the genetic bases of these and other neurological diseases.

Differential editing

The thousands of neurons profiled in the original study by Lake and colleagues [19] enabled testing of editing differences at individual sites across neuronal sub-types and cortical regions. As the data was too sparse to allow post-hoc pair-wise testing between all neuronal sub-types and cortical regions (96 groups), we compared sites by gross neuronal phenotype (inhibitory and excitatory) across the six cortical regions (12 groups). Despite the abundance of cells from the superior temporal cortex (BA41),

463 the majority of differential editing derived from enhanced editing in frontal cortical neurons (BA8 and BA10). This finding is
464 supported by the greater global editing rates observed in both neuronal sub-types in the frontal cortex compared to other regions.
465 Interestingly, BA17 and BA41 house the primary visual and auditory cortices, respectively, and exhibited lower global editing
466 rates. If higher editing rates are assumed to dampen neuronal activity in both inhibitory and excitatory cell types, then these
467 regional differences may support the prevailing understanding of the functional organization of the cortex, wherein sensory
468 regions provide strong excitatory input (corresponding to reduced editing); and the frontal cortex is primarily inhibitory, acting
469 to increase the salience of selected perceptual inputs [53]. Further work to integrate functional imaging/electrocorticography
470 data with single-cell sequencing, will greatly assist in elaborating the relationship between the sequence and composition of the
471 neuronal transcriptome, neuronal location, and function.
472
473 Lastly, our cell and region-specific differential editing results allow high-resolution examination of epigenetic processes implicated
474 in neuropsychiatric disease. Schizophrenia (SCZ) has been associated with relatively higher global RNA editing rates than in
475 healthy controls [7], whereas the opposite is reported for autism spectrum disorder (ASD) [6]. Enhanced editing in the frontal
476 cortex of four out of six DEd sites in KCNIP4 in our data is particularly interesting given the significant hypo-editing in of 11
477 proximal sites in introns 1 and 2 of this transcript in the dorso-lateral prefrontal cortex of SCZ patients [7]. KCNIP4 regulates the
478 function of the potassium channel Kv4, which propagates action potentials in inhibitory (GABAergic) neurons [54,55]. The two
479 DEd KCNIP4 sites in our data that are implicated in SCZ, appear preferentially edited in inhibitory neurons of the superior frontal
480 cortex; and excitatory neurons of the frontal and temporal cortex, respectively. Similarly, attractin-like protein 1 is the most hypo-
481 edited transcript in the anterior cingulate cortex of SCZ patients, and was found to be preferentially edited in excitatory neurons
482 of fronto-temporal regions in our healthy brain data. These findings, together with our identification of region and cell-type
483 specific DEd sites in synaptic transmembrane proteins (LRRTM4) and non-coding RNAs (FTX and SNRPN) that are hypo-edited in
484 SCZ, suggest spatially defined and cell-type specific regulation of these transcripts, that may yield highly specific targets for future
485 therapeutic interventions.
486
487 Whereas fewer sites implicated in ASD were DEd in our results, both conditions appear to involve differential editing at the 15q11
488 region containing snoRNA host genes. The site 15:24983610, in SNHG14, is hypo-edited in SCZ, hyper-edited in ASD, and found
489 here to be preferentially edited in inhibitory neurons of the frontal (BA8) and temporal (BA41) cortex. SNHG14 contains a cluster
490 of editing sites, 5 of which are DEd in SCZ specifically. The proximal gene SNRPN runs in to SNHG14 and is also hypo-edited in SCZ.
491 Interestingly, RBFOX family proteins regulate the production of snoRNAs at 15q11 [56–58]. The ASD 'hub gene' RBFOX1 [59] is
492 hypo-edited in the disorder [6], and found to be DEd at two sites in the frontal and temporal cortex in our results. Together these
493 findings indicate that the 15q11 region is a nexus of RNA editing-related activity whose transcription is a proxy for global RNA
494 editing rates; and impinges on neurotransmitter receptor function (serotonin receptors in particular) [60,61]. The extensive
495 differential editing reported in 15q11 in this study, between neuronal sub-types and cortical regions; and in two prominent
496 neuropsychiatric disorders, makes this region a compelling subject for further investigation. It is also intriguing to consider the
497 potential involvement of RNA editing dysregulation in neurological conditions associated with structural variants in the 15q11
498 locus [62,63].
499
500 In conclusion, we integrated RNA editing signals in thousands of individual neurons with known editing sites in the healthy and
501 diseased brain. These sites reflect cell-type and region-specific epigenetic processes that may be negated by genetic mutations
502 in rare neurological diseases, and provide targets for gene therapeutic interventions in neuropsychiatric disease. The single-cell
503 level resolution achieved in this study revealed thousands of previously undetected editing sites with genomic features similar to
504 documented sites, which replicate in independent single-cell studies, but which are likely subsumed in studies of bulk RNA. This

work adds exciting new dimensions to our understanding of post-transcriptional regulation in the central nervous system, and a comprehensive reference for forthcoming investigation of editing in single-cell samples from neuropsychiatric patient cohorts.

COMPETING INTERESTS

The authors declare that they have no competing interests.

DATA AVAILABILITY

Raw RNA sequencing data for phs000834 is available upon application from the Database of Genotypes and Phenotypes (dbGAP). Raw RNA sequencing data for GSE67835 is available at the Sequence Read Archive (SRA). Summary data is available in Supplementary tables, and scripts for reproducing results in this manuscript from summary tables are freely available at github.com/bahlolab/brain_scRNAed.

REFERENCES

1. Slotkin W, Nishikura K. Adenosine-to-inosine RNA editing and human disease. Genome Med. Genome Medicine; 2013;5:105–18.

2. Hwang T, Park C-K, Leung AKL, Gao Y, Hyde TM, Kleinman JE, et al. Dynamic regulation of RNA editing in human brain development and disease. Nat Neurosci. Nature Publishing Group; 2016;19:1093–9.

3. Srivastava PK, Bagnati M, Delahaye-Duriez A, Ko J-H, Rotival M, Langley SR, et al. Genome-wide analysis of differential RNA editing in epilepsy. Genome Res. 2017;27:440–50.

4. Silvestris DA, Picardi E, Cesarini V, Fosso B, Mangraviti N, Massimi L, et al. Dynamic inosinome profiles reveal novel patient stratification and gender-specific differences in glioblastoma. Genome Biol. 2019;20:33.

5. Lyddon R, Dwork AJ, Keddache M, Siever LJ, Dracheva S. Serotonin 2c receptor RNA editing in major depression and suicide. World J Biol Psychiatry. 2013;14:590–601.

6. Tran SS, Jun H-I, Bahn JH, Azghadi A, Ramaswami G, Van Nostrand EL, et al. Widespread RNA editing dysregulation in brains from autistic individuals. Nat Neurosci. 2019;22:25–36.

7. Breen MS, CommonMind Consortium, Dobbyn A, Li Q, Roussos P, Hoffman GE, et al. Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia [Internet]. Nature Neuroscience. 2019. p. 1402–12. Available from: http://dx.doi.org/10.1038/s41593-019-0463-7

8. Li JB, Levanon EY, Yoon J-K, Aach J, Xie B, LeProust E, et al. Genome-Wide Identification of Human RNA Editing Sites by Parallel DNA Capturing and Sequencing. Science. 2009;324:1210–3.

9. Higuchi M, Maas S, Single FN, Hartner J, Rozov A, Burnashev N, et al. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature. Nature Publishing Group; 2000;406:78–81.

10. Livingston JH, Lin J-P, Dale RC, Gill D, Brogan P, Munnich A, et al. A type I interferon signature identifies bilateral striatal necrosis due to mutations in ADAR1. J Med Genet. 2014;51:76–82.

11. Gallo A, Vukic D, Michalík D, O'Connell MA, Keegan LP. ADAR RNA editing in human disease; more to it than meets the I. Hum Genet. Springer Berlin Heidelberg; 2017;136:1265–78.

12. Moore S, Alsop E, Lorenzini I, Starr A, Rabichow BE, Mendez E, et al. ADAR2 mislocalization and widespread RNA editing aberrations in C9orf72-mediated ALS/FTD [Internet]. Acta Neuropathologica. 2019. p. 49–65. Available from: http://dx.doi.org/10.1007/s00401-019-01999-w

13. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published

549 genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47:D1005–12.

550 14. Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, et al. Dynamic landscape and regulation of RNA editing in
551 mammals. Nature. Nature Publishing Group; 2017;550:249–54.

552 15. Nishikura K. Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat Rev Mol Cell Biol. Nature
553 Publishing Group; 2006;7:919–31.

554 16. Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, et al. Non-optimal codon usage affects expression, structure and function
555 of clock protein FRQ. Nature. 2013;495:111–5.

556 17. Picardi E, Horner DS, Pesole G. Single-cell transcriptomics reveals specific RNA editing signatures in the human brain. RNA.
557 2017;23:860–5.

558 18. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human brain transcriptome diversity at the
559 single cell level. Proc Natl Acad Sci U S A. National Acad Sciences; 2015;112:7285–90.

560 19. Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA
561 sequencing of the human brain. Science. 2016;352:1586–90.

562 20. Ramaswami G, Li JB. RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic Acids Res. 2013;42:D109–13.

563 21. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update
564 2006. Nucleic Acids Res. 2006;34:D590–8.

565 22. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res.
566 2002;12:996–1006.

567 23. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation.
568 Nucleic Acids Res. 2001;29:308–11.

569 24. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations
570 and supporting evidence. Nucleic Acids Res. 2018;46:D1062–7.

571 25. Hudson WH, Ortlund EA. The structure, function and evolution of proteins that bind DNA and RNA. Nat Rev Mol Cell Biol.
572 2014;15:749–60.

573 26. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. Nat Rev Mol Cell Biol. 2016;17:83–96.

574 27. Jorjani H, Kehr S, Jedlinski DJ, Gumienny R, Hertel J, Stadler PF, et al. An updated human snoRNAome. Nucleic Acids Res.
575 2016;44:5068–82.

576 28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics.
577 2012;29:15–21.

578 29. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic
579 features. Bioinformatics. 2014;30:923–30.

580 30. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with
581 Bioconductor. F1000Res. 2016;5:2122.

582 31. McCarthy DJ, Campbell KR, Lun ATL, Wills QF. Scater: pre-processing, quality control, normalization and visualization of
583 single-cell RNA-seq data in R. Bioinformatics. 2017;33:1179–86.

584 32. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high
585 confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1–33.

586 33. Zheng X, Gogarten SM, Lawrence M, Stilp A, Conomos MP, Weir BS, et al. SeqArray-a storage-efficient high-performance
587 data format for WGS variant calls. Bioinformatics. 2017;33:2251–7.

588 34. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

589 35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools.
590 Bioinformatics. 2009;25:2078–9.

591 36. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol.
592 2016;17:122.

593 37. R Core Team. R: A Language and Environment for Statistical Computing. 2014; Available from: R-project.org

594 38. Wickham H, Francois R, Henry L, Müller K, Others. dplyr: A grammar of data manipulation. R package version 0 4. 2015;3.

595 39. Wickham H, Henry L. Tidyr: Easily tidy data with spread() and gather() functions. R package version 0 6. 2017;1.

596 40. Weston S, Microsoft Corporation. doParallel: Foreach parallel adaptor for the parallel package. R package version [Internet].
597 2018; Available from: https://CRAN.R-project.org/package=doParallel

598 41. Wickham H. multidplyr: A Multi-Process "dplyr" Backend [Internet]. 2019. Available from:
599 https://github.com/tidyverse/multidplyr

600 42. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer; 2016.

601 43. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties.
602 Bioinformatics. Oxford University Press; 2017;33:2938–40.

603 44. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinformatics.
604 2014;30:2811–2.

605 45. Robinson D, Hayes A. Broom: Convert statistical analysis objects into tidy tibbles. R package version 0 5 0 https://CRAN R-
606 project org/package= broom. 2018;

607 46. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing
608 and microarray studies. Nucleic Acids Res. 2015;43:e47–e47.

609 47. Mangiafico S. rcompanion: Functions to support extension education program evaluation. R package version. 2017;1.

610 48. Quinones-Valdez G, Tran SS, Jun H-I, Bahn JH, Yang E-W, Zhan L, et al. Regulation of RNA editing by RNA-binding proteins in
611 human cells. Commun Biol. 2019;2:19.

612 49. Ullah MI, Ahmad A, Raza SI, Amar A, Ali A, Bhatti A, et al. In silico analysis of SIGMAR1 variant (rs4879809) segregating in a
613 consanguineous Pakistani family showing amyotrophic lateral sclerosis without frontotemporal lobar dementia. Neurogenetics.
614 2015;16:299–306.

615 50. Hodge RD, Bakken TE, Miller JA, Smith KA, Barkan ER, Graybuck LT, et al. Conserved cell types with divergent features in
616 human versus mouse cortex. Nature. 2019;573:61–8.

617 51. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol
618 Cell Biol. 2007;8:209–20.

619 52. Feng Y, Sansam CL, Singh M, Emeson RB. Altered RNA editing in mice lacking ADAR2 autoregulation. Mol Cell Biol.
620 2006;26:480–8.

621 53. Hu Y, Chen X, Gu H, Yang Y. Resting-state glutamate and GABA concentrations predict task-induced deactivation in the
622 default mode network. J Neurosci. 2013;33:18566–73.

623 54. Markram H, Toledo-Rodriguez M, Wang Y, Gupta A, Silberberg G, Wu C. Interneurons of the neocortical inhibitory system
624 [Internet]. Nature Reviews Neuroscience. 2004. p. 793–807. Available from: http://dx.doi.org/10.1038/nrn1519

625 55. Cattenoz PB, Taft RJ, Westhof E, Mattick JS. Transcriptome-wide identification of A > I RNA editing sites by inosine specific
626 cleavage. RNA. 2013;19:257–70.

627 56. Jacko M, Weyn-Vanhentenryck SM, Smerdon JW, Yan R, Feng H, Williams DJ, et al. Rbfox Splicing Factors Promote Neuronal
628 Maturation and Axon Initial Segment Assembly. Neuron. 2018;97:853–68.e6.

629 57. Coulson RL, Powell WT, Yasui DH, Dileep G, Resnick J, LaSalle JM. Prader–Willi locus Snord116 RNA processing requires an
630 active endogenous allele and neuron-specific splicing by Rbfox3/NeuN. Hum Mol Genet. Narnia; 2018;27:4051–60.

631 58. Conboy JG. Developmental regulation of RNA processing by Rbfox proteins. Wiley Interdiscip Rev RNA [Internet]. 2017;8.
632 Available from: http://dx.doi.org/10.1002/wrna.1398

633 59. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals
634 convergent molecular pathology. Nature. 2011;474:380–4.

635 60. Kishore S. The snoRNA HBII-52 Regulates Alternative Splicing of the Serotonin Receptor 2C [Internet]. Science. 2006. p. 230–
636 2. Available from: http://dx.doi.org/10.1126/science.1118265

637 61. Vitali P, Basyuk E, Le Meur E, Bertrand E, Muscatelli F, Cavaillé J, et al. ADAR2-mediated editing of RNA substrates in the
638 nucleolus is inhibited by C/D small nucleolar RNAs. J Cell Biol. 2005;169:745–53.

639 62. Jønch AE, Douard E, Moreau C, Van Dijck A, Passeggeri M, Kooy F, et al. Estimating the effect size of the 15Q11.2 BP1-BP2
640 deletion and its contribution to neurodevelopmental symptoms: recommendations for practice. J Med Genet. 2019;56:701–10.

641 63. Ulfarsson MO, Walters GB, Gustafsson O, Steinberg S, Silva A, Doyle OM, et al. 15q11.2 CNV affects cognitive, structural and
642 functional correlates of dyslexia and dyscalculia. Transl Psychiatry. 2017;7:e1109.

643

644 **SUPPLEMENTARY FIGURE LEGENDS**
645
646 **SFigure 1. Independent clustering of neuronal nuclear transcriptomes separates inhibitory from excitatory**
647 **neurons.** UMAP plot displaying independent clustering of neuronal transcriptomes based on gene expression. Each
648 point represents a neuronal nuclear transcriptome, coloured by cell type as assigned by Lake and colleagues (Lake
649 et al. 2016). Inhibitory neurons (In) area at left, and excitatory neurons (Ex) at right.
650
651 **SFigure 2. Read mapping statistics for neuronal nuclear transcriptomes, grouped by cortical area, and neuronal**
652 **sub-group.** The average number of uniquely mapped reads was 3 million. The mean number of multi-mapping reads
653 was 780,535. When quantifying gene expression, a fractional assignment method was used for multi-mapping reads
654 (see methods). BA8 showed significantly more uniquely mapped reads than other cortical regions; and neuronal sub-
655 groups In7 and In8 had significantly lower numbers of uniquely mapped reads; whereas Ex2 and In6 had higher
656 numbers.
657
658 **SFigure 3. Novel editing sites are detected in fewer cells than previously reported sites.** Across all three site
659 contexts (Alu repeat, repetitive non-Alu, and non-repetitive), novel sites are detected in significantly fewer neurons.
660 These lower-prevalence sites may be lost in the ensemble averaging inherent in bulk RNA sequencing, in which
661 previously-reported sites were identified (RADAR database).
662
663 **SFigure 4. Reference codon context for 467 edited sites within coding domains, faceted by position of edited**
664 **adenosine (A).**
665
666 **SFigure 5. Distribution of minor allele frequency (aka frequency of inosine; FI) for 40,861 putative editing sites.**
667 Bars are colored according to previously reported genomic context (RADAR database; (Ramaswami and Li 2013)).
668 Note that 'novel' sites include numerous site located in e.g. Alu repeats, but which are not present in the RADAR
669 database. See Figure 1c & d.
670
671 **SFigure 6. Minimal effect of removal of SNHG14 editing site cluster from editing proportion calculations.** A)
672 Correlation in global editing proportions including (x axis) and excluding (y axis) the editing site cluster in SNHG14
673 cluster. B) Correlation between editing proportion calculated without the SNHG14 editing site cluster (x axis), and
674 expression of representative SNORD115 cluster snoRNA HBII-52. For comparison with Figure 4c.
675

676 **SUPPLEMENTARY TABLE TITLEs**
677
678 Please see **scRNAed_Dictionaries.xlsx** for data dictionaries describing the data in each of the following tables.
679
680 **STAble 1.** Metadata for 3055 neurons published by Lake et al, *Science* 2016. All neurons were isolated from the left
681 hemisphere of a single donor (female, 51 y/o). Neuronal nuclear transcriptomes were sequenced using SMART-seq
682 chemistry.
683
684 **STable 2.** Metadata for 116 neurons isolated from four unrelated donors, published by Darmanis et al, *PNAS* 2015.
685 Neuronal nuclear transcriptomes were sequenced using SMART-seq chemistry.
686
687 **STable 3.** Genomic context and intersect with published studies for 40,851 sites detected across 3,055 neuronal
688 transciptomes from Lake et al, *Science,* 2016.
689
690 **STable 4.** Linear model summaries for test of global editing index (GEI) between neurons from i) diffcortical cortical
691 regions and ii) different neuronal sub-types.
692
693 **STable 5.** Summary for linear models testing the relationship between expression of non-edited transcripts, and
694 global editing index (GEI).
695
696 **STable 6.** Enriched gene ontology terms linked to genes whose transcription is significantly correlated with global
697 editing index (GEI).
698
699 **STable 7.** Variant effect predictions for edited sites in coding domains, and reported editing enzymes.
700
701 **STable 8.** Clinical variants associated with genomic variants underlying edited sites, and phenotype and cortical
702 region of neurons exhibiting the edited site.
703
704 **STable 9.** Gene ontology terms enriched among genes containing differentially edited sites.
705
706 **STable 10.** Differentially edited sites between neuronal subtypes and cortical areas, intersected with differentially
707 edited sites implicated in schizophrenia (Breen et al, *Nat Neurosci*, 2019) and autism spectrum disorder (Tran et al,
708 *Nat Neurosci*, 2018).
709
710
711
712
713
714