Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia

Michael S. Breen1,2,3*, Amanda Dobbyn2,4,5, Qin Li6, Panos Roussos1,2,5,7,8, Gabriel E. Hoffman2,5,9, Eli Stahl1,2,5,9,10, Andrew Chess2,8,9,11, Pamela Sklar8,9,31, Jin Billy Li6, Bernie Devlin12, Joseph D. Buxbaum1,2,3,8,13* and CommonMind Consortium32

RNA editing critically regulates neurodevelopment and normal neuronal function. The global landscape of RNA editing was surveyed across 364 schizophrenia cases and 383 control postmortem brain samples from the CommonMind Consortium, comprising two regions: dorsolateral prefrontal cortex and anterior cingulate cortex. In schizophrenia, RNA editing sites in genes encoding AMPA-type glutamate receptors and postsynaptic density proteins were less edited, whereas those encoding translation initiation machinery were edited more. These sites replicate between brain regions, map to 3′-untranslated regions and intronic regions, share common sequence motifs and overlap with binding sites for RNA-binding proteins crucial for neurodevelopment. These findings cross-validate in hundreds of non-overlapping dorsolateral prefrontal cortex samples. Furthermore, ~30% of RNA editing sites associate with cis-regulatory variants (editing quantitative trait loci or edQTLs). Fine-mapping edQTLs with schizophrenia risk loci revealed co-localization of eleven edQTLs with six loci. The findings demonstrate widespread altered RNA editing in schizophrenia and its genetic regulation, and suggest a causal and mechanistic role of RNA editing in schizophrenia neuropathology.

Schizophrenia (SCZ) is a severe psychiatric disorder affecting ~0.7% of adults and is characterized by abnormalities in thought and cognition. Although the onset of SCZ typically does not occur until late adolescence or early adulthood, strong support from clinical and epidemiological studies suggests that SCZ reflects a disturbance of neurodevelopment. There is clear and consistent evidence that SCZ is largely a genetic disorder. Large-scale mapping of genetic risk variants has identified multiple rare copy number variants, several rare single nucleotide variants and >100 common genetic loci, the latter exerting small polygenetic effects on disease risk. This observation of a highly polygenic architecture has been widely replicated. However, the role of sequence variation arising as a result of post-transcriptional RNA modifications, such as RNA editing, remains largely unexplored.

RNA editing is a modification of double-stranded pre-mRNA that introduces changes in mRNA through insertions, deletions or substitutions of nucleotides, and hence can lead to alterations in protein function. Adenosine-to-inosine (A-to-I) editing is the most common form of RNA editing, affecting most human genes, and is highly prevalent in the brain. These base-specific changes to RNA result from site-specific deamination of nucleotides catalyzed by the enzymes known as adenosine deaminases acting on RNA (ADAR), whereby a genetically encoded adenosine is edited into an inosine, which is read by the cellular machinery as a guanosine. Editing sites in coding regions can be conserved across species and are commonly located in genes involved in neuronal function. RNA editing has been reported to influence permeability of ion channels, excitatory responses and other neuronal signaling functions. These sites have been shown to be tightly and dynamically regulated throughout pre- and postnatal human cortical development. Abrupt RNA editing has also been reported in several neurological disorders, including major depression, Alzheimer’s disease and amyotrophic lateral sclerosis.

In SCZ, the role of RNA editing in serotonin and glutamate receptors has drawn significant attention largely due to the serotonergic and glutamatergic hypotheses of mood disorders. To this end, RNA editing research in SCZ has focused on targeted approaches of the serotonin 2C receptor (5-HT2CR) and two classes of ionotropic glutamate receptors, AMPA and kainate receptors. Consequently, there is no consensus on the type of editing or on how pervasive altered RNA editing is in the brain of SCZ patients. Moreover, the underlying cis-acting genetic variants, which are associated with edQTLs in the brain, and whether these variants are also implicated in disease risk, also remain poorly understood.

1Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 2Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 3Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 4The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 5Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 6Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. 7Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA. 8Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 9ICahn Institute for Data Science and Genomic Technologies, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 10Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. 11Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. 12Department of Cell Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 13Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. 14Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 31Deceased: Pamela Sklar 32A full list of authors and affiliations appears at the end of the paper. *e-mail: michael.breen@mssm.edu; joseph.buxbaum@mssm.edu
The primary goal of the current investigation was to clarify the relevance of RNA editing in SCZ pathophysiology using an unbiased, genome-wide approach applied to a large cohort of SCZ cases and control samples generated from the CommonMind Consortium (CMC), which is orders of magnitude larger than prior RNA editing studies. Two brain regions implicated in neurodevelopment and SCZ neuropathology were examined, including the dorsolateral prefrontal cortex (DLPFC; Brodmann areas 9 and 46) and anterior cingulate cortex (ACC). Results from this cohort were then reproduced in a separate, non-overlapping DLPFC cohort generated through the National Institute of Mental Health (NIMH) Human Brain Collection Core (HBCC). By applying a multi-step analytic framework and including genome-wide characterization of common genetic variation, a resource was generated for the genetics of RNA editing in the brain. This resource was used to identify: (1) genes and RNA editing sites with significant differences in edQTLs between subjects with SCZ and control subjects; (2) coordinated editing (co-editing) of RNA editing sites implicated in SCZ; and (3) specific effects on RNA editing of genetic variants previously implicated in disease risk. In doing so, these findings substantially refine our understanding of the RNA editing-mediated mechanism involved in the neurobiology of SCZ.

Results

Discovery and validation samples. To quantify RNA editing events, RNA-sequencing (RNA-seq) data were taken from post-mortem brain tissue collected and generated on behalf of the CMC. Two brain regions, including the ACC (SCZ = 225, controls = 245) and the DLPFC (SCZ = 254, controls = 286) were investigated, and together these samples served as the discovery cohort (see Supplementary Fig. 1). These samples were also genotyped on the Illumina Infinium HumanOmniExpressExome array. In parallel, a completely separate, non-overlapping cohort was also taken, consisting of postmortem DLPFC tissue (SCZ = 100, controls = 217) collected and generated on behalf of the NIMH HBCC. This second resource served as a validation cohort to cross-validate the discovery of SCZ-related editing events (Fig. 1).

Overall RNA editing levels in SCZ. Overall RNA editing levels were computed for each sample and were defined as the percentage of edited nucleotides at all known editing sites (see Methods). Higher levels of overall RNA editing in SCZ cases were observed compared with controls in the ACC and DLPFC (P = 0.0001, P = 7.2 × 10−5, respectively) (Fig. 2a). Approximately 10% of the variation in overall RNA editing levels was explained by ADAR1 (P = 2.2 × 10−14) and ADAR2 expression explained ~3% of the variation in overall RNA editing (P = 1.2 × 10−4) (Fig. 2b,c). ADAR3 expression had no significant effect on overall editing levels (P = 0.10) (Fig. 2d), although recently it demonstrated a negative association with overall editing when measured across several brain regions. Marked increases in overall editing levels were observed within definite genic regions, specifically 3′-untranslated region (3′-UTR) and intergenic regions in SCZ, which replicated across the ACC and DLPFC (see Supplementary Fig. 2a). Moreover, as previous research has quantified RNA editing levels explicitly in serotoninergic and glutamatergic receptors, overall editing levels were computed in serotoninergic and glutamatergic receptor activity genes using assumed defined gene sets (GO:009589 and GO:0008066, respectively). Higher levels of overall editing were found in glutamatergic receptors in SCZ cases relative to controls in the ACC (P = 0.0001) and DLPFC (P = 2.2 × 10−5), whereas no significant differences were found in the levels of overall editing in serotoninergic receptors (see Supplementary Fig. 2b). Expression of ADAR1 and ADAR2 was also notably higher in SCZ compared with control samples (see Supplementary Fig. 2c–e). Collectively, these observations, apart from the expression of ADAR2, were reproduced in the independent DLPFC validation cohort in the present study, and collectively highlight higher overall RNA editing levels in SCZ, primarily within 3′-UTR and intergenic regions, as well as within genes encoding glutamatergic receptors.

To rule out the possibility that these reproducible differences in overall RNA editing levels may be driven by medication effects, overall editing levels were examined in postmortem DLPFC tissue derived from an RNA-seq study of 34 rhesus macaque monkeys treated with high doses of haloperidol (10 mg kg−1 d−1), low doses of haloperidol (4 mg kg−1 d−1), clozapine (5.2 mg kg−1 d−1) and vehicle. No associations were found between overall RNA editing levels with medication or dosage (see Supplementary Fig. 3), indicating that antipsychotic treatments probably do not have a strong effect on the amount of overall RNA editing observed in SCZ cases.

Discovery of altered RNA editing sites in SCZ. To identify RNA editing sites associated with SCZ, a compendium of high-quality and high-confidence RNA editing sites was assembled by imposing a series of detection-based thresholds (see Methods). After a thorough quality control, a high-confidence set of 11,242 RNA editing sites was identified in the ACC and 7,594 sites in the DLPFC, with no systematic differences in the mapping, base quality and read coverage between SCZ and control samples. A significant fraction of these RNA editing sites replicated across brain regions (n = 6,999, odds ratio (OR) = 21.05, P < 2.0 x 10−10). A large fraction of the sites were located in Alu repeat elements, mapped to 3′-UTR regions and were enriched for A-to-I conversions (see Supplementary Fig. 4).
Fig. 2 | Overall RNA editing profiles. a, Overall RNA editing levels across the CMC ACC (n\textsubscript{control} = 245, n\textsubscript{SCZ} = 225), DLPFC (n\textsubscript{control} = 286, n\textsubscript{SCZ} = 254) and HBCC DLPFC (n\textsubscript{control} = 217, n\textsubscript{SCZ} = 100). A two-sided Mann–Whitney U-test with continuity correction was used to test for differences between diagnostic groups. Whisker box plots show median, lower and upper quartiles, and whiskers represent the minimum and maximum of the data. Associations between expression levels of ADAR1 (b), ADAR2 (c) and ADAR3 (d) (quantified as the number of RPKMs) and overall editing levels across all available ACC and DLPFC samples (including CMC and HBCC data). These concordance analyses were made across all samples (n\textsubscript{control} = 735, n\textsubscript{SCZ} = 579) because the ACC and DLPFC showed highly collinear relationships. R
squared values were calculated using robust linear regressions on overall editing levels and logarithmic transformed RPKM values.

After this curation of editing events, differential RNA editing analysis was carried out. It is probable that genome-wide RNA editing events, similar to gene expression, may be influenced by differences in biological and technical factors. To this end, a linear mixed-effect model was applied to quantify the total amount of RNA editing variance explained by various biological and technical factors, which collectively displayed little influence on editing profiles, with individual age having the largest genome-wide effect, and explained a median 0.79% of the observed variability (see Supplementary Fig. 5a). These factors, however, explained a much higher amount of median variability in matching gene expression profiles than observed RNA editing profiles (see Supplementary Fig. 5b). Subsequently, differential editing analysis covarying for individual age, RNA integrity number (RIN), postmortem interval (PMI), sample site and sex identified 182 sites in the ACC and 194 sites in the DLPFC significantly associated with SCZ (adjusted \(P < 0.05\) (Fig. 3a and see Supplementary Table 1a–c). Among the top-ranked sites were those encoding for the genes ATRN1, AKAP5 and RPS20 in the ACC and KCNIP4, VPS41 and ZNF140 in the DLPFC. A high degree of concordance was observed between the altered RNA editing sites in the ACC and DLPFC (\(R^2 = 0.59\)) and a significant overlap of differentially edited sites replicated between brain regions (\(n_{\text{rep}} = 29, \text{OR} = 12.6, P = 9.6 \times 10^{-29}\) (Fig. 3b)). Differentially edited sites were also enriched in genes found to be highly expressed in the ACC (\(n_{\text{rep}} = 42, \text{OR} = 5.4, P = 2.0 \times 10^{-11}\)) and DLPFC (\(n_{\text{rep}} = 54, \text{OR} = 8.6, P = 2.7 \times 10^{-21}\)) and that most genes with differential RNA editing sites did not display differential gene expression (see Supplementary Fig. 6a–f). Moderate, yet notable, correlations were also observed between RNA editing levels and gene expression, implying RNA editing as a possible post-transcriptional mechanism for the regulation of gene expression (see Supplementary Fig. 6g–i).

Validation of altered RNA editing sites in SCZ. Next, we asked whether these differentially editing patterns in SCZ replicate within the independent DLPFC validation cohort. These samples underwent matching quality-control metrics to identify a collection of highly confident RNA editing events, as noted above. A total of 15,000 RNA editing events were detected across these validation samples and a significant fraction of sites was also detected in the ACC (74%, \(n_{\text{rep}} = 8,354, \text{OR} = 4.01, P = 5.6 \times 10^{-25}\)) and DLPFC (87%, \(n_{\text{rep}} = 6,659, \text{OR} = 7.73, P = 4.67 \times 10^{-44}\) discovery samples (see Supplementary Table 2). Differential RNA editing analysis was carried out on these independent samples as previously described and 137 sites (75%) were detected in the ACC and 165 sites (85%) in the DLPFC. To assess replication, first the concordance was measured between directionality of change in editing rates for all RNA editing sites identified in the ACC and for all the DLPFC discovery samples, relative to these independent DLPFC validation samples. High levels of concordance were observed across all RNA editing sites in both the ACC and the DLPFC (\(R = 0.12, R^2 = 0.13\), respectively) (Fig. 3c,d). Subsequently, two prediction models were built based on differentially edited sites from (1) the DLPFC and (2) the ACC discovery samples using regularized regression models, and their performance evaluated to predict class labels (that is, distinguish between SCZ and control samples) on withheld DLPFC validation samples. Classification accuracies were reported as area under the receiver operative curve on withheld DLPFC samples. When distinguishing between SCZ and control samples, classification accuracies reach 78% and 72% on withheld, independent DLPFC samples if using differentially edited sites derived from DLPFC and ACC discovery samples, respectively (ridge regression outperformed other methods; see Supplementary Fig. 7). Overall, these results suggest a moderate level of cross-validation of SCZ-related editing events across brain regions and independent cohorts.

Characterization of differentially edited sites. Differentially edited sites derived from selection and validation samples were comprehensively annotated. Although most differentially edited sites map to 3’-UTR regions across brain regions and cohorts, a moderate depletion was observed when adjusting for the total number of non-differentially edited sites in 3’-UTRs for each brain region and cohort (Fig. 3e). Functional enrichment analysis revealed that under-edited sites consistently mapped to postsynaptic density genes, as well as genes encoding kainate and glutamate receptor activity, and over-edited sites mapped to genes implicated in protein translation and mitochondrially related terms (Fig. 3f,g). The study also examined whether these differentially edited sites map to genes with specific developmental expression profiles using gene expression data from the BrainSpan Project, and it was found that differentially edited sites in SCZ consistently mapped to genes that are predominantly postnatally biased in expression (see Supplementary Fig. 8). These genes were found to peak in brain expression during developmental windows in young and middle adulthood, when SCZ often becomes clinically recognizable. Moreover, a substantial fraction of our editing sites (\(n = 612\)) was also previously found to have increasing rates of editing throughout brain development, including 11 that are notably over-edited and 3 that are under-edited in SCZ (see Supplementary Table 1d).
As these sites share several sequence and functional features, it was explored whether differential editing sites may share a common sequence motif potentially important for editome recognition (20±nucleotides (nt) centered on target A). Consistent enrichment was found for a 10-nt motif (CGGGATTACA) in the region adjacent to most differential and non-differential editing sites located within the ACC and DLPFC (HBCC) regions. The motif was enriched in the region adjacent to most differential and non-differential editing sites located within the ACC and DLPFC (HBCC) regions.
Articles

Genes enriched with differential RNA editing in SCZ. The present study examined whether any genes contained an enrichment of differentially edited sites beyond what could be expected by chance. As expected, gene length functions as a correlate of the total number of RNA editing sites per gene (see Supplementary Fig. 10). Therefore, over-representation of differential RNA editing sites within each gene was computed by setting a rotating background specific to the total number of known RNA editing events for a particular gene, to systematically correct for gene length (see Supplementary Table 5). Genes harboring an important fraction of under-edited sites in SCZ primarily mapped to intronic regions (Fig. 4a,b), whereas genes harboring 3′-UTR sites were over-edited in SCZ (Fig. 4b–e). Three genes, including KCNIP4, HOOK3 and MRPS16, displayed enrichment for altered editing sites across the ACC, DLPFC and independent validation DLPFC cohort (Fig. 4c–e). KCNIP4 harbored 13, unique, differentially edited sites spread over its first and second introns, which were predominantly under-edited in SCZ compared with control samples. KCNIP4 is a member of the voltage-gated, potassium channel-interacting proteins, and has been shown to interact with presenilins and modulate pacemaker neurons in the heart. KCNIP4 has been suggested to play a role in SCZ23,24. NOVA1 is essential for neuronal toxicity.

Identification and characterization of brain cis-edQTLs. Whole-genome genotype data were available for ACC and DLPFC samples used in the discovery cohort of the present study and were imputed using standard techniques, as previously described. Genotype data were used to detect SNPs that have an effect on edQTLs. RNA editing levels from European-ancestry samples (ACC n = 360; DLPFC n = 421) were adjusted to fit a standard normal distribution and to reduce systematic sources of variation. Adjusted editing levels were then filtered to impute SNP genotypes, covarying for individual age, sample site and gender, PMI, RIN and diagnosis, using an additive linear model implemented in MatrixEQT. To identify genetic variants that could explain the variability of RNA editing, first association tests were run between editing levels and genotypes by restricting the variant search space to only those within the same gene as each editing site, and an abundance of low P values was found (see Supplementary Fig. 13a). Subsequently, this assumption was relaxed to define a broader window and 188,778 cis-edQTLs were identified (that is, SNP-editing pairs ± 100 kilobases (kb) of a site) in the ACC and 156,865 cis-edQTLs in the DLPFC at a genome-wide false discovery rate (FDR) < 5% (Fig. 6a). A total of 3,224 editing sites in the ACC and 2,500 editing sites in the DLPFC have edQTLs. Many of the edQTLs for the same site were highly correlated, due to linkage disequilibrium, and 70.9% of edQTL SNPs (edSNPs) in the ACC and 68.9% in the DLPFC predicted editing of more than one site. A high level of concordance was observed for the effect sizes (beta values) of edQTLs between the ACC and DLPFC (see Supplementary Fig. 14). Notably, edQTLs tend to be present for editing sites with greater variance in editing levels (see Supplementary Fig. 15). Each max-edQTL (defined as the most important edSNP per site, if any) meeting a genome-wide significance threshold was located close to its associated editing site and acting in cis (5 kb ± nt) (Fig. 6b and see Supplementary Fig. 15). It was reasoned that, due to the propensity for edQTLs to be located close to their associated editing site, they should also influence additional editing sites nearby. This reasoning was strengthened by the observation that editing levels of editing sites within the same gene are more closely correlated than editing levels of editing sites in different genes (see Supplementary Fig. 16).

Max-edQTLs in the ACC and DLPFC were enriched within genetic elements and non-coding RNAs, particularly within intronic regions, whereas the corresponding editing sites were also enriched in intronic regions and depleted from 3′-UTR regions (Fig. 6c). Max-edQTL edSNPs were also examined for a tissue-specific enhancer that specifies using data from the FANTOM project across 40 different human tissues. The edSNPs in the ACC and DLPFC were strongly

ARTICLES

NATURE NEUROSCIENCE

in 3′-UTR regions (see Supplementary Fig. 9 and Supplementary Table 3). Notably, this short sequence has been reported to occur frequently within non-coding regions and is also found to overlap fragments of Alu repeat elements. Subsequently, it was examined whether differentially edited sites found to share this sequence motif also mapped to any known human RNA-binding protein (RBP)-binding sites (30 ± nt centered on target A). A large fraction of these sites (>61% in each brain region) importantly coincided with binding sites specific for RBP serine/arginine (S/R)-rich splicing factor 5 (SR5F5) (see Supplementary Table 4). It is interesting that this protein is associated with pyruvate carboxylase deficiency, a disorder that is associated with developmental delay and recurrent seizures. Other significant RBP-binding sites included additional members of the S/R-rich family of pre-mRNA splicing factors, such as SR5F2 and SR5F3, as well as CUGBP, an RBP found to mediate neuronal toxicity.

Co-editing networks associated with SCZ. Discrete groups of coordinately edited (co-edited) sites were identified and tested for association with SCZ using an unbiased network approach. A total of five co-editing modules was detected in each brain region and displayed a near one-to-one mapping between the ACC and the DLPFC (Fig. 5a), indicating highly similar co-editing network topology. Modules were assessed for over-representation of differential RNA editing sites and two modules were identified in the ACC (M1a and M4a) and two in the DLPFC (M1d and M4d) (Fig. 5b). Module eigengene (ME) values for these modules elucidated higher levels of editing in modules M1a and M1d and lower levels in modules M4a and M4d in SCZ compared with control subjects (Fig. 5c). Functional annotation of over-edited modules M1a and M1d revealed strong enrichment for regulation of translation and translation initiation, whereas under-edited modules M4a and M4d were enriched for AMPA-glutamate and ionotropic receptors (Fig. 5d). Cell-type enrichment analysis revealed that modules M1a and M1d were enriched for pyramidal neurons, whereas modules M4a and M4d were enriched for interneurons (Fig. 5e). Notably, these findings were also reproduced in our independent DLPFC validation cohort (see Supplementary Table 6 and Supplementary Fig. 11). Moreover, M1a and M1d were positively associated with ADAR1 and ADAR2 expression and modules M4a and M4d were negatively associated with ADAR2 expression (see Supplementary Fig. 12). On closer inspection, several sites located within modules M4a and M4d mapped to non-synonymous sites in the genes NOVA1, UNC80, GRIA2, GRIA3, GRIA4, GRIK2 and ANKD36, and these sites were predominantly under-edited in SCZ compared with control samples (Fig. 5f). Several of these sites, particularly the Q/R and R/G sites in GRIA2, are well documented as fully edited sites under normal conditions, whereby loss of editing in these sites leads to enhanced Ca2+ permeability and cellular dysfunction, and this has been suggested to play a role in SCZ. NOVA1 is essential for normal postnatal motor function and regulates alternative splicing of multiple inhibitory synaptic targets. NOVA1 has been reported to be dysregulated at the gene level in independent SCZ postmortem brain samples and RNA editing in NOVA1 has been shown to influence protein stability, but has yet to be associated with SCZ.
enriched for brain-specific enhancer sequences, more so than any other tissue (see Supplementary Fig. 17). An important fraction of max-eQTL edSNPs was replicated between the ACC (62%) and the DLPPC (70%) (n = 34,367, Z-score = 17,443, P = 0.0009). Among the most important associations identified in both brain regions were those in genes H2AFV and PNMAL1, where the edSNP is located immediately upstream of the RNA editing site (Fig. 6d–g). In both cases, the alternative allele is unable to pair with the opposite base within the double-stranded RNA hairpin, introducing two consecutive mismatches in the local RNA secondary structure.
In addition, edSNPs were examined for association with gene expression levels by calculating the overlap between max-edQTLs and previously computed maximum expression QTL summary statistics derived from the ACC and the DLPFC. A total of 29,335 edSNPs (54.4%) in the ACC were also associated with variation in gene expression, for which 31.3% were associated with a gene and one or more editing sites located within the same gene (for example, SNP, is associated with Gene, and one or more editing sites located within Gene). Similarly, a total of 27,133 edSNPs in the DLPFC (55.6%) were also associated with gene expression variation, for which 30.1% were associated with a gene and one or more editing sites within the same gene.

The edQTL signatures co-localize with SCZ GWAS associations. It has been previously shown that a substantial proportion of SCZ GWAS associations (~20%) may be mediated by differential gene expression regulation. RNA editing may represent an additional mechanism that contributes to SCZ risk. In the present study, the edQTL resource was used to identify RNA editing sites that potentially alter SCZ risk. Of the 108 SCZ GWAS loci reported previously, 14 harbor edQTLs for one or more RNA editing sites identified in either ACC or DLPFC. However, the presence of an edQTL within a GWAS locus does not imply disease causality. Therefore, the presence of an edQTL within a GWAS locus does not imply disease causality. Therefore, colorectal was implemented, a Bayesian approach that integrates over statistics for all variants within a specified locus and estimates posterior probabilities of co-localization between two sets of association signatures, to identify RNA editing sites likely to contribute to SCZ etiology. Coloc2 was applied to the ACC and DLPFC edQTL data together with summary statistics for the 108, genome-wide, important, schizophrenia GWAS loci. Evidence was found for co-localization (posterior probability >0.5) of ACC edQTL and GWAS signatures at four loci comprising unique edQTLs and of DLPFC edQTLs and GWAS signatures at four loci comprising seven unique edQTLs (see Supplementary Table 7). Two of these loci are co-localized in both ACC and DLPFC; therefore, a total of six GWAS associations, representing approximately 5% of all genome-wide significant loci, are potentially mediated by aberrant RNA editing (see Supplementary Fig. 18). Of the six GWAS loci harboring SCZ-associated cis-edQTLs, these findings include genes NGEF and ARL6IP4, which replicate between brain regions, as well as PCCB and RP11-890B15.3, which are unique to the ACC, and genes ENSA and UNC80.

| Gene | ACC | DLPFC |
|------|-----|-------|
| GRIA4 | M1a | M1d |
| GRIA2 | M1a | M1d |
| GRIK2 | M1a | M1d |
| NOVA1 | M1a | M1d |
| RP11-890B15.3 | M1a | M1d |
| ... | ... | ... |

In addition, edSNPs were examined for association with gene expression levels by calculating the overlap between max-edQTLs and previously computed maximum expression QTL summary statistics derived from the ACC and the DLPFC. A total of 29,335 edSNPs (54.4%) in the ACC were also associated with variation in gene expression, for which 31.3% were associated with a gene and one or more editing sites located within the same gene (for example, SNP, is associated with Gene, and one or more editing sites located within Gene). Similarly, a total of 27,133 edSNPs in the DLPFC (55.6%) were also associated with gene expression variation, for which 30.1% were associated with a gene and one or more editing sites within the same gene. 

The edQTL signatures co-localize with SCZ GWAS associations. It has been previously shown that a substantial proportion of SCZ GWAS associations (~20%) may be mediated by differential gene expression regulation. RNA editing may represent an additional mechanism that contributes to SCZ risk. In the present study, the edQTL resource was used to identify RNA editing sites that potentially alter SCZ risk. Of the 108 SCZ GWAS loci reported previously, 14 harbor edQTLs for one or more RNA editing sites identified in either ACC or DLPFC. However, the presence of an edQTL within a GWAS locus does not imply disease causality. Therefore colorectal was implemented, a Bayesian approach that integrates over statistics for all variants within a specified locus and estimates posterior probabilities of co-localization between two sets of association signatures, to identify RNA editing sites likely to contribute to SCZ etiology. Coloc2 was applied to the ACC and DLPFC edQTL data together with summary statistics for the 108, genome-wide, important, schizophrenia GWAS loci. Evidence was found for co-localization (posterior probability >0.5) of ACC edQTL and GWAS signatures at four loci comprising unique edQTLs and of DLPFC edQTLs and GWAS signatures at four loci comprising seven unique edQTLs (see Supplementary Table 7). Two of these loci are co-localized in both ACC and DLPFC; therefore, a total of six GWAS associations, representing approximately 5% of all genome-wide significant loci, are potentially mediated by aberrant RNA editing (see Supplementary Fig. 18). Of the six GWAS loci harboring SCZ-associated cis-edQTLs, these findings include genes NGEF and ARL6IP4, which replicate between brain regions, as well as PCCB and RP11-890B15.3, which are unique to the ACC, and genes ENSA and UNC80.

In addition, edSNPs were examined for association with gene expression levels by calculating the overlap between max-edQTLs and previously computed maximum expression QTL summary statistics derived from the ACC and the DLPFC. A total of 29,335 edSNPs (54.4%) in the ACC were also associated with variation in gene expression, for which 31.3% were associated with a gene and one or more editing sites located within the same gene (for example, SNP, is associated with Gene, and one or more editing sites located within Gene). Similarly, a total of 27,133 edSNPs in the DLPFC (55.6%) were also associated with gene expression variation, for which 30.1% were associated with a gene and one or more editing sites within the same gene. 

The edQTL signatures co-localize with SCZ GWAS associations. It has been previously shown that a substantial proportion of SCZ GWAS associations (~20%) may be mediated by differential gene expression regulation. RNA editing may represent an additional mechanism that contributes to SCZ risk. In the present study, the edQTL resource was used to identify RNA editing sites that potentially alter SCZ risk. Of the 108 SCZ GWAS loci reported previously, 14 harbor edQTLs for one or more RNA editing sites identified in either ACC or DLPFC. However, the presence of an edQTL within a GWAS locus does not imply disease causality. Therefore colorectal was implemented, a Bayesian approach that integrates over statistics for all variants within a specified locus and estimates posterior probabilities of co-localization between two sets of association signatures, to identify RNA editing sites likely to contribute to SCZ etiology. Coloc2 was applied to the ACC and DLPFC edQTL data together with summary statistics for the 108, genome-wide, important, schizophrenia GWAS loci. Evidence was found for co-localization (posterior probability >0.5) of ACC edQTL and GWAS signatures at four loci comprising unique edQTLs and of DLPFC edQTLs and GWAS signatures at four loci comprising seven unique edQTLs (see Supplementary Table 7). Two of these loci are co-localized in both ACC and DLPFC; therefore, a total of six GWAS associations, representing approximately 5% of all genome-wide significant loci, are potentially mediated by aberrant RNA editing (see Supplementary Fig. 18). Of the six GWAS loci harboring SCZ-associated cis-edQTLs, these findings include genes NGEF and ARL6IP4, which replicate between brain regions, as well as PCCB and RP11-890B15.3, which are unique to the ACC, and genes ENSA and UNC80.
and DGKI, which are unique to the DLPFC; co-localization of the DGKI locus is highlighted in Fig. 7.

Discussion
The recent expansion of RNA-seq data-sets has led to the identification of a huge number of RNA editing events, which affect most human genes and are highly prevalent in the brain. Many such sites are commonly located in genes involved in neuronal maintenance, and aberrant editing events have been associated with various neurological disorders. However, it has yet to be understood how pervasive RNA editing events are in the brain of SCZ patients and what genetic forces are guiding the regulation of these events. The ACC and DLPFC have been shown to play an important role in neurodevelopment, and have been implicated in the pathophysiology of SCZ.

Lower levels of RNA editing were associated with postsynaptic K+-channel functions. The first intron of KCNIP4 in neurons. The first intron of KCNIP4 is involved in the regulation of the potassium channel Kv4, which is a significant contributor to action potential activity in neurons. The first intron of KCNIP4 is involved in alternative splicing events, leading to Var IV of KCNIP4, which has been found to disrupt this current through failure to properly interact with presenilins, a component of the γ-secretase complex. It is plausible that RNA editing may influence splice-site choice in KCNIP4, leading to aberrant neuronal functioning through modulation of Kv4 channel functions.

Higher levels of RNA editing were observed in genes that are essential for mitochondrial protein translation. One of these genes harboring over-edited sites in SCZ was the RNA-binding motif protein 8A (RBM8A), which has been shown to control mRNA stability and splicing, and translation, and is located in the 1q21.1 copy-number-variation-associated autism spectrum disorder, SCZ and...
Articles

Articles NATuRE NeuRoSCIeNCe

Affect neuronal activity, including synaptic connection, axon formation, and cell metabolism and membrane excitability. The diversity of cell metabolism and membrane excitability allows for exploitation by differential pharmacology, creating in-roads into new targeted pharmacological interventions.

In conclusion, the present study reveals dynamic aspects of RNA editing in human brain tissue covering hundreds of SCZ cases and control samples, including two brain regions and two large primary cohorts used for discovery and validation. Strong reproducible evidence was identified for widespread dysregulation of RNA editing in SCZ, including under-editing of glutamate receptor activity and post-synaptic density genes, which show pyramidal neuronal cell-type specificity, as well as over-editing in genes involved in regulation of translation and translation initiation, which are specific to interneuronal cell types. Moreover, a large portion of RNA editing sites was characterized as being involved in cis-edQTLs in human brain tissue and GWAS–edQTL co-localization analysis was further performed, which identified co-localization of 11 edQTLs with 6 GWAS loci. This result supports a causal role of RNA editing in risk for SCZ. Although these results shed new light into the mechanisms underlying the neuropathophysiology of SCZ, additional molecular studies of aberrant RNA editing sites identified in the present study and their molecular mechanisms are required to fully appreciate their functional importance for SCZ neurobiology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0463-7.

Received: 2 December 2018; Accepted: 9 July 2019; Published online: 27 August 2019

References

1. McGrath, J., Saha, S., Chant, D. & Welham, J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiol. Rev. 30, 67–76 (2008).
2. Owen, M., Sawa, A. & Mortensen, P. Schizophrenia. Lancet 384, 86–97 (2016).
3. Kirov, G. CNVs in neuropsychiatric disorders. Hum. Mol. Genet. 24, R45–R49 (2015).
4. Xu, B. et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat. Genet. 44, 1365–1369 (2012).
5. Takata, A. et al. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. Neuron 82, 773–780 (2014).
6. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427 (2014).
7. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. 19, 1442–1452 (2016).
8. Meier, S. M. et al. High loading of polygenic risk in cases with chronic schizophrenia. Mol. Psychiatry 21, 969–974 (2016).

Fig. 7 | Coloc2 fine-mapping analysis. GWAS and edQTL summary statistics (beta, standard error) for SNPs within each GWAS locus were used as input for coloc2. a Loci (x-axis) according to imputed genotype dosages (y-axis) and maximum of adjusted RNA editing levels (y-axis) were visualized. b GWAS (a) and edQTL (b) signal is on chromosome 7 DGLI locus (PPH4 = 0.99). This specific co-localization event is specific to the DLPFC. Linkage disequilibrium estimates are colored with respect to the GWAS lead SNP (rs3735025) and coded as a heatmap from dark blue (r ≥ 0.99) to red (0.8 ≥ r > 0.2). Notably, genes NGEG and ARIL6IP4 replicated between brain regions. The edSNPs and editing sites for NGEG are located within 3′-UTR regions and enhancer elements. NGEG is predominantly brain expressed, particularly during early development, and shows substantial homology with the Dbl family, which are implicated in human cognitive function. The editing site in ARIL6IP4 (also known as splicing factor SRp25) causes a non-synonymous amino acid substitution (K/R) and affects a basic region in the protein that has not been ascribed a specific function. GWAS–edQTL co-localization was also identified for ENSA, a gene that belongs to a highly conserved CAMP-regulated phosphoprotein family, and is considered an endogenous regulator of ATP-sensitive potassium (KATP) channels, which rest at the intersection of cell metabolism and membrane excitability. The diversity of KATP channel properties allows for exploitation by differential pharmacology, creating in-roads into new targeted pharmacological interventions.

microcephaly. Moreover, several independent reports indicate mitochondrial dysfunction in schizophrenia, which can severely affect neuronal activity, including synaptic connection, axon formation and neuronal plasticity. A future concerted approach to sites encoding mitochondrial genes will provide a more complete understanding of how editing in these genes impacts SCZ neurobiology.

It was determined that edQTLs are widespread in brain tissue and a substantial portion replicates between two brain regions. Approximately 30% of all RNA editing sites were associated with one or more nearby cis-regulatory variants. It is expected that the genomics of cis-edQTLs and their RNA editing sites align with context-specific regulation of editing, as indicated through overlap of edSNPs with regulatory elements, such as tissue-specific enhancers (see Supplementary Fig. 17), and mapping of RNA editing sites on genes, which are predominately postnatally biased in neocortical gene expression (see Supplementary Fig. 8). Moreover, six GWAS loci demonstrate co-localization with edQTLs and show moderate effect sizes (β = 0.82 ± 0.26). Notably, genes NGEG and ARIL6IP4 replicated between brain regions. The edSNPs and editing sites for NGEG are located within 3′-UTR regions and enhancer elements.

NGEG is predominantly brain expressed, particularly during early development, and shows substantial homology with the Dbl family, which are implicated in human cognitive function. The editing site in ARIL6IP4 (also known as splicing factor SRp25) causes a non-synonymous amino acid substitution (K/R) and affects a basic region in the protein that has not been ascribed a specific function. GWAS–edQTL co-localization was also identified for ENSA, a gene that belongs to a highly conserved CAMP-regulated phosphoprotein family, and is considered an endogenous regulator of ATP-sensitive potassium (KATP) channels, which rest at the intersection of cell metabolism and membrane excitability. The diversity of KATP channel properties allows for exploitation by differential pharmacology, creating in-roads into new targeted pharmacological interventions.

In conclusion, the present study reveals dynamic aspects of RNA editing in human brain tissue covering hundreds of SCZ cases and control samples, including two brain regions and two large primary cohorts used for discovery and validation. Strong reproducible evidence was identified for widespread dysregulation of RNA editing in SCZ, including under-editing of glutamate receptor activity and post-synaptic density genes, which show pyramidal neuronal cell-type specificity, as well as over-editing in genes involved in regulation of translation and translation initiation, which are specific to interneuronal cell types. Moreover, a large portion of RNA editing sites was characterized as being involved in cis-edQTLs in human brain tissue and GWAS–edQTL co-localization analysis was further performed, which identified co-localization of 11 edQTLs with 6 GWAS loci. This result supports a causal role of RNA editing in risk for SCZ. Although these results shed new light into the mechanisms underlying the neuropathophysiology of SCZ, additional molecular studies of aberrant RNA editing sites identified in the present study and their molecular mechanisms are required to fully appreciate their functional importance for SCZ neurobiology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0463-7.

Received: 2 December 2018; Accepted: 9 July 2019; Published online: 27 August 2019

References

1. McGrath, J., Saha, S., Chant, D. & Welham, J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiol. Rev. 30, 67–76 (2008).
2. Owen, M., Sawa, A. & Mortensen, P. Schizophrenia. Lancet 384, 86–97 (2016).
3. Kirov, G. CNVs in neuropsychiatric disorders. Hum. Mol. Genet. 24, R45–R49 (2015).
4. Xu, B. et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat. Genet. 44, 1365–1369 (2012).
5. Takata, A. et al. Loss-of-function variants in schizophrenia risk and SETD1A as a candidate susceptibility gene. Neuron 82, 773–780 (2014).
6. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427 (2014).
7. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. 19, 1442–1452 (2016).
8. Meier, S. M. et al. High loading of polygenic risk in cases with chronic schizophrenia. Mol. Psychiatry 21, 969–974 (2016).
9. Behm, M. & Öhman, M. RNA editing: a contributor to neuronal dynamics in the mammalian brain. Trends Genet. 32, 165–175 (2016).

10. Tan, M. et al. Dynamic landscape and regulation of RNA editing in mammals. Nature 550, 249–254 (2017).

11. Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. Nat. Rev. Mol. Cell Biol. 17, 83–96 (2016).

12. Rosenthal, J. J. & Seeberg, P. H. A-to-I RNA editing: effects on proteins key to neural excitability. Neuron 74, 432–439 (2012).

13. Rula, E. Y. et al. Developmental modulation of GABAA receptor function by RNA editing. J. Neurosci. 28, 6196–6201 (2008).

14. Krestel, H. E. et al. A genetic switch for epilepsy in adult mice. J. Neurosci. 24, 10568–10578 (2004).

15. Wahlstedt, H., Daniel, C., Estérom, M. & Öhman, M. Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. Genome Res. 19, 978–986 (2009).

16. Lydon, R., Dworak, J. A., Keddache, M., Siever, L. J. & Dracheva, S. Serotonin 2c receptor RNA editing in major depression and suicide. World J. Biol. Psychiatry 14, 590–601 (2013).

17. Gaisler-Salomon, I. et al. Hippocampus-specific deficiency in RNA editing of GLA2 in Alzheimer’s disease. Neurobiol. Aging 35, 1785–1791 (2014).

18. Kwak, S. & Kawahara, Y. D eficient RNA editing of GLU2 and neuronal death in amyotrophic lateral sclerosis. J. Mol. Med. 83, 110–120 (2005).

19. Herrick-Davis, K., Grinde, E. & Niswander, C. M. Serotonin 5-HT2C receptor RNA editing alters receptor basal activity: implications for serotonergic signal transduction. J. Pharmacol. Exp. Ther. 309, 1711–1717 (2004).

20. Marion, S., Weiner, D. M. & Caron, M. G. RNA editing induces variation in desensitization and trafficking of 5-hydroxytryptamine 2c receptor isoforms. J. Biol. Chem. 279, 2945–2954 (2004).

21. Sodhi, M. S., Burnet, P. W., Makoff, A. J., Kerwin, R. W. & Harrison, P. J. RNA editing of the 5-HT 2C receptor is reduced in schizophrenia. Mol. Psychiatry 6, 373–379 (2001).

22. Dracheva, S. et al. Increased serotonin 2C receptor mRNA editing: a possible risk factor for suicide. Mol. Psychiatry 13, 1001–1010 (2008).

23. Sommer, B., Köhler, M., Sprengel, R. & Seeburg, P. H. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. Cell 67, 11–19 (1991).

24. Kubota-Sakashita, M., Iwamoto, K., Bundo, M. & Kato, T. I. A role of ADAR2 and RNA editing of glutamate receptors in mood disorders and schizophrenia. Mol. Brain 7, 5 (2014).

25. Lomeli, H. et al. Control of kinetic properties of AMPA receptor channels by RNA editing. Neuron 277, 7286–7296 (2007).

26. Sullivan, P. F. et al. Genomewide association for schizophrenia in the CATIE Neurogenetics Consortium. Brain 17, 12–23 (2014).

27. Akrouh, A., Halcomb, S. E., Nichols, C. G. & Sala-Rabanal, M. Molecular biology of KATP channels and implications for health and disease. IUBMB Life 69, 971–978 (2009).

Acknowledgements Data were generated as part of the CommonMind Consortium, supported by funding from Takeda Pharmaceuticals Company Ltd, F. Hoffman-La Roche Ltd and NIH grant nos. R01MH088542, R01MH093735, P50MH066392, P50MH080405, RO1MH097276, RO1-MH-075916, P01MH06981, P50MH084053S1, R37MH057881, AG02219, AG015138, MH06692, RO1MH110921, RO1MH10677, RO1MH98977 and U01MH033922, and contract no. HHSN272201300031C through the NIH’s Intramural Research Program. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer’s Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories, and the NIH HBCCC. CMC leadership comprises: P. Roussos, J. D. Burxbaum, A. Chess, S. Akbarian, V. Haroutunian (Icahn School of Medicine at Mount Sinai), B. Devlin, D. Lewis (University of Pittsburgh), R. Gur, C.-G. Hahn (University of Pennsylvania), E. Domenici (University of Trento), M. A. Peters, S. Marenco, S. Sieberts (Sage Bionetworks), T. Lehner, G. Senthil, S. Marenco and B. K. Lipska (NIH). DLPEF RNA-seq data, which formed the basis of the validation cohort, were provided by the NIH HBCCC. Rhesus macaque tissue was provided by S. Hembry through the Stanley Medical Research Institute for Funding for Non-Human Primate Research, and funded by NIH grant no. R01MH074313.

Author contributions J.D.B., P.S., I.J.B.L. and M.S.B contributed to experimental design, study design and formulating the research question. J.D.B. and P.S. contributed to the funding of this work. M.S.B., A.D. and Q.L. contributed to data analysis. P.R., G.E.H., E.S., A.C., P.S., B.D. and J.D.B. contributed to leadership and supervision of various aspects of this work. M.S.B. and J.D.B. contributed to writing the manuscript, and all authors contributed in completing the final version.

Competing interests The authors declare no competing interests.

Additional information Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0463-7. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.S.B. or J.D.B. Peer review information: Nature Neuroscience thanks T. Kato, J. Yang and the other anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
CommonMind Consortium

Schahram Akbarian1,16, Jaroslav Bendl1,5, Kristen Brennand1,5, Leanne Brown1,16, Andrew Browne3, Alexander Charney1,5, Lizette Couto1,16, Greg Crawford17, Olivia Devillers1,16, Enrico Domenici18, Michele Filosi18, Elie Flatow1,16, Nancy Francoeur5, John F. Fullard1,2,16, Sergio Espeso Gil1,16, Kiran Girdhar1,5, Attila Gulyás-Kovács11, Raquel Gur19,20, Chang-Gyu Hahn21, Vahram Haroutunian1,5,7,22, Mads Engel Hauberg8,23, Laura Huckins1,5, Rivky Jacobov1,16, Yan Jiang1,16, Jessica S. Johnson1,5, Bibi Kassim1,16, Yungil Kim1,5, Lambertus Klei16, Robin Kramer24, Mario Lauria25, Thomas Lehner26, David A. Lewis27, Barbara K. Lipska24, Stefano Marenco24, Lara M. Mangravite28, Kelsey Montgomery28, Royce Park1,16, Thanneer Malai Perumal28, Mette A. Peters28, Chaggai Rosenbluh1, Douglas M. Ruderfer27,29, Geetha Senthil26, Hardik R. Shah2,9, Solveig K. Sieberts28, Laura Sloofman1,5, Lingyun Song30, Patrick Sullivan31, Roberto Visintainer18, Jiebiao Wang32, Ying-Chih Wang2,9, Jennifer Wiseman1,16, Eva Xia11, Wen Zhang1,5 and Elizabeth Zharovsky1,16

16Division of Psychiatric Epigenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 17Center for Genomic & Computational Biology, Duke University, Durham, NC, USA. 18Laboratory of Neurogenomic Biomarkers, Centre for Integrative Biology, University of Trento, Trento, Italy. 19Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 20Neuropsychiatry Section, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. 21Neuropsychiatric Signaling Program, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. 22Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 23Department of Biomedicine, Aarhus University, Aarhus, Denmark. 24Human Brain Collection Core, National Institutes of Health, National Institute of Mental Health, Bethesda, MD, USA. 25Department of Mathematics, University of Trento, Trento, Italy. 26Department of Biomedicine, University of Aarhus, Aarhus, Denmark. 27Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. 28Sage Bionetworks, Seattle, WA, USA. 29Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. 30Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. 31Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. 32Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA, USA.
Methods

Identification of RNA editing sites from human RNA-seq data. RNA-seq data generated from the human postmortem ACC (n\textsubscript{sample}=245, n\textsubscript{cell}=225) and DLPCF (n\textsubscript{sample}=286, n\textsubscript{cell}=294) were obtained through the CMC (number of uniquely mapped reads, 31,626,646 and 12,539,625). Additional RNA-seq data from human postmortem DLPCF samples (n\textsubscript{sample}=217, n\textsubscript{cell}=100) were obtained through the NIH HBCC (number of uniquely mapped reads, 105,426,854 ± 10,0006,209). All fastq files were mapped to human reference genome hg19 using STAR v2.4.049 and the following parameters were optimized: chimSegmentMin = 15; chimJunctionOverhangMin = 15; outSAMstrandField = none;SAMSOFT=1; and STAR-F2-3. For each sample, this produced a coordinate-sorted bam file of mapped paired-end reads, including those spanning splice junctions. Known RNA edited sites were curated using the publicly available database, Rigorously Annotated Database of A-to-I editing sites and were used for downstream analysis. For each sample, this produced a coordinate-sorted bam file of mapped paired-end reads, including those spanning splice junctions. Known RNA edited sites were curated using the publicly available database, Rigorously Annotated Database of A-to-I editing sites and were used for downstream analysis. To identify a collection of high-quality and high-confidence sites, a series of detection-based thresholds were placed for each brain region and cohort, separately: (1) the minimum base quality of 20; (2) minimum mapping quality of 20 (that is, probability that a read is aligned to multiple locations); (3) probability of misalignment = 0.01 (that is, 99% probability that a read is correctly aligned in the genome); (4) minimum read coverage per edited site to be 20—the identification of RNA editing sites has previously been reported to be prone to these biases, so it is likely that making this parameter more lenient would increase the number of falsely predicted editing events; (5) all known SNPs present in the SNP database (except SNPs of molecular type cDNA) and those within the 1000 Genomes Project were also removed; and (6) an editing site was required to be present in at least 80% of all samples and, subsequently, must have no more than 20% missing values per sample. The resulting RNA editing data frames for the ACC, HBCC, and DLPCF samples contained 8.3% and 9.8% missing data, respectively, and the data frame for HBCC DLPCF samples contained 7.6% missing data. All missing values were imputed using predictive mean matching method in the mice R package51, using 5 multiple imputations and 30 iterations. The resulting sets of sites identified from these RNA-seq data were subsequently referred to as ‘known’ RNA editing sites and were used for downstream analysis.

No statistical methods were used to predetermine sample sizes; however, the sample sizes of the present study are the largest to be reported. All samples used in the present study were from participants in two large studies of schizophrenia in the USA that were donated their brains on death. Data collection and analysis were not performed blind to the conditions of the experiments. No animals or data points were excluded from the analyses for any reason.

Identification of RNA editing sites from macaque RNA-seq data. To examine whether drug treatment effects were responsible for overall RNA editing levels observed in SCZ, overall editing derived from RNA-seq study of DLPCF tissue from rhesus macaque monkeys was computed. Antipsychotic administration, performed blind to the conditions of the experiments. No animals or data points were excluded from the analyses for any reason.

Differential RNA editing analysis. It is possible that RNA editing levels, similar to those observed in gene expression studies, are influenced by a number of biological and technical factors. Therefore, before differential RNA editing analysis, the variance for each site was partitioned into the variance attributable to each variable using a linear mixed model implemented in the R package variancePartition51. Under this framework, categorical variables (that is, disease status, biological sex) are modeled as random effects and continuous variables (that is, individual age, PMI) are modeled as fixed effects. Each site was considered separately and the results for all sites were aggregated afterwards. This approach enabled the rational inclusion of leading covariates into the downstream analysis, which may ultimately have an influence on differential RNA editing analysis. Subsequently, to identify sites with differential RNA editing levels between SCZ and control samples, a linear model was implemented through the limma R package22 covarying for the possible influence of individual age, RIN, PMI, sample site and sex. Significance values were adjusted for multiple testing using the Benjamini and Hochberg method to control the FDR. Sites passing a multiple-test, corrected P value < 0.05 were labeled significant. To further demarcate the confounding effects of antipsychotic medications on editing levels, SCZ-related sites were examined that were also conserved in the rhesus macaque samples, as noted above. Of these sites, 24 unique sites were identified with sufficient coverage (>10 reads per site) across all rhesus macaque DLPCF samples that were also not differentially edited in at least one SCZ brain region. Using these sites, a series of pairwise comparisons was made using a linear model implemented through limma to compute the change in editing rates associated with 5.2 mg clozapine, 0.14 mg haloperidol and 4 mg haloperidol relative to vehicle treatment (see Supplementary Table 1e). Subsequently, the concordance between antipsychotic-induced changes in editing rates in macaques and SCZ-related changes in humans was calculated using a robust linear regression and non-parametric approach. To identify candidate SCZ sites and antipsychotic medications for this subset of conserved RNA editing events (see Supplementary Table 1e). In addition, additional measures were taken to ensure that the landscape of RNA editing in SCZ was not confounded by differences in cellular composition. RNA-seq fastq files of adult human brain single cells were downloaded from the Gene Expression Omnibus database using the accession number GSE67835. Raw RNA-seq files were aligned to the human reference (hg19) using STAR alignment with default paired-end parameters. The present study identified 35 unique sites with sufficient coverage (>5 reads per site) across at least 70% or more of all adult human brain samples from at least one SCZ brain region. A lower coverage threshold was implemented due to RNA-seq coverage being orders of magnitude lower than the bulk postmortem tissue RNA-seq samples. Next, no more than 80% of missing values were allowed for each individual cell type, thereby yielding a total of 181 cells dissociated from adult brain cortex, which were in the present study’s analysis, including oligodendrocytes (n = 16), oligodendrocyte precursors (n = 7), astrocytes (n = 39) and neurons (n = 119). To identify changes in editing rates associated with cell type differences, levels of RNA editing were compared between neuronal and non-neuronal cell types using a linear model, as described above. Similarly, the concordance between neuronal-related differences in editing rates was evaluated relative to SCZ-related changes in humans using a robust linear regression and no notable associations for this subset of RNA editing events were found (see Supplementary Table 1f).

Supervised class prediction methods. To assess cross-validation of the SCZ-related sites, two prediction models were built using the differentially edited sites in (1) the DLPCF and (2) the ACC derived from the CMC (here referred to as training set) to predict case/control status (that is, SCZ cases from control samples) from different DLPCF RNA-seq data derived from the HBCC (here referred to as test set). Regularized regression models, including ElasticNet, Lasso and Ridge Regression, were fitted using the glmnet R package. The penalty parameter lambda ($\lambda$) was estimated using tenfold cross-validation on each training set using the caret package in R, and ultimately set to lambda.min, the value of $\lambda$ that yields minimum mean cross-validated error of each regularisation model. Once optimal $\lambda$ was identified, they were applied to RNA editing levels from the test set using the predict() function, which calculates the predicted logarithm of the odds of diagnostic status. Subsequently, the area under the receiver operative curve analysis was performed using the pROC package in R. Classification accuracies were reported as area under the curve on test samples to assess the precision of the models.

Identification of enriched sequence motifs and RBP sites. Previous studies suggest that RNA editing events are mediated by RBPs that recognize specific
BrainSpan developmental gene-set enrichment analysis. BrainSpan developmental RNA-seq data (www.brainspan.org) were summarized to GENCODE10, and gene-level RNA-seq reads per kilobase of transcript per million mapped reads (RPKM) values were extracted across 528 samples. Only the neocortical regions were used in the analysis—dorsolateral prefrontal cortex, ventrolateral prefrontal cortex, medial prefrontal cortex, orbitofrontal cortex, primary motor cortex, primary somatosensory cortex, primary association cortex, inferior parietal cortex, superior temporal cortex, inferior temporal cortex and primary visual cortex. A RIN ≥ 7 was used as a filtering threshold from subsequent analysis. Genes were defined as if they were present at an RPKM of 0.5 in 80% of the samples, from at least one neocortical region at one major temporal epoch, resulting in 22,141 transcripts across 299 high-quality samples ranging from postconception week 8 to 40 years of age. Finally, expression values were log-transformed (log, RPKM+1)).

Further information on research design is available in the Nature Research Reporting Summary linked to this article.
Data availability

The CMC investigators are committed to the release of data and analysis results, with the anticipation that data sharing in a rapid and transparent manner will speed the pace of research to the benefit of the greater research community. Data and analytical results generated through the CMC are available through the CMC Knowledge Portal: https://doi.org/10.7303/syn2759792.

URLs

HBCC: https://www.nimh.nih.gov/research/research-conducted-at-nimh/research-areas/research-support-services/hbcc/index.shtml

CMC: http://www.synapse.org/CMC

Code availability

Code for identifying RNA editing sites and quantifying RNA editing ratios are provided in the public repository: https://github.com/BreenMS/RNAediting

Differential RNA editing, co-editing network analyses and edQTL analysis used standard software packages.

References

49. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
50. Ramaswami, G. & Li, J. B. RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic Acids Res. 42, D109–D113 (2014).
51. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
52. Bueren, S. V. & Groothuis-Oudshoorn, K. Mice: multivariate imputation by chained equations in R. J. Stat. Softw. 45, 1–68 (2010).
53. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics 25, 1841–1842 (2009).
54. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics 17, 483 (2016).
55. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47–e47 (2015).
56. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22 (2010).
57. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 12–77 (2011).
58. Bailey, T. L., Williams, N., Misleb, C. & Li, W. W. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 34, W369–W373 (2006).
59. Paz, I., Kosti, I., Ares, M. Jr, Cline, M. & Mandel-Gutfreund, Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Res. 42, W361–W367 (2014).
60. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).
61. Chen, Jing et al. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 37, W305–W311 (2009).
62. Zeisel, A. et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 347, 1138–1142 (2015).
63. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353–1358 (2012).
64. Lizio, M. et al. Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals. Nucleic Acids Res. 4, D737–D743 (2016).
65. Gel, B. et al. Regioner: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics 32, 289–291 (2015).
66. Dobbyn, A. et al. Landscape of conditional eQTL in dorsolateral prefrontal cortex and co-localization with schizophrenia GWAS. Am. J. Hum. Genet. 102, 1169–1184 (2018).
Corresponding author(s): Michael S. Breen and Joseph D. Buxbaum
Last updated by author(s): July 3, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| Item                                                                 | Confirmed |
|----------------------------------------------------------------------|-----------|
| n/a The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement | Yes       |
| A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly | Yes       |
| The statistical test(s) used AND whether they are one- or two-sided | Yes       |
| Only common tests should be described solely by name; describe more complex techniques in the Methods section. | Yes       |
| A description of all covariates tested | Yes       |
| A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons | Yes       |
| A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) | Yes       |
| For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted | Yes       |
| Give P values as exact values whenever suitable. | Yes       |
| For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings | Yes       |
| For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes | Yes       |
| Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated | Yes       |

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

We leveraged RNA-sequencing data from post-mortem brain tissue collected and generated on behalf of the CommonMind Consortium (CMC). Two brain regions, including the anterior cingulate cortex (SCZ=225, Controls=245) and the dorsolateral prefrontal cortex (DLPFC; SCZ=254, Controls=286) were investigated. These samples were also genotyped on the Illumina Infinium HumanOmniExpressExome array. In parallel, we also leveraged a completely separate, non-overlapping cohort consisting of RNA-sequencing data from the post-mortem DLPFC tissue (SCZ=100, Controls=217) collected and generated on behalf of National Institute of Mental Health (NIMH) Human Brain Collection Core (HBCC). All data are downloadable through the CMC synapse.org panel.

Data analysis

All RNA-sequencing fastq files were mapped to human reference genome hg19 using STAR version 2.4.01. Known RNA-edited sites were curated using the publicly available database, Rigorously Annotated Database of A-to-I RNA editing (RADAR). Nucleotide coordinates for these well-documented editing sites were then used to extract reads from each sample using a customized perl script (https://github.com/BreenMS/RNAediting) and the samtools mpileup function. Other tools and respective versions are as follows: STAR version 2.5.0; mice version 3.5.0; track layer version 2.13; ToppGene Suite (http://toppgene.chcmc.org); variancePartition version 1.12.08; limma version 3.38.2; glmnet version 2.0-16; pROC version 1.13.0; MEME version 5.0.3; RBMap version 5.12; WGCNA version 1.66; matrixEQTL version 2.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The CommonMind investigators are committed to the release of data and analysis results with the anticipation that data sharing in a rapid and transparent manner will speed the pace of research to the benefit of the greater research community. Data and analytical results generated through the CommonMind Consortium are available through the CommonMind Consortium Knowledge Portal: http://dx.doi.org/10.7303/syn27259792

URLs.
Human Brain Collection Core (HBCC): https://www.nimh.nih.gov/research/research-conducted-at-nimh/research-areas/research-support-services/hbcc/index.shtml
CommonMind Consortium (CMC): http://www.synapse.org/CMC
Code availability. Code for identifying RNA editing sites and quantifying RNA editing ratios are provided in the public repository: https://github.com/BreenMS/RNAediting
Differential RNA editing, co-editing network analyses and edQTL analysis used standard software packages.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
We leveraged RNA-sequencing data from post-mortem brain tissue collected and generated on behalf of the CommonMind Consortium (CMC). Two brain regions, including the anterior cingulate cortex (SCZ=225, Controls=245) and the dorsolateral prefrontal cortex (DLPFC; SCZ=254, Controls=286) were investigated. These samples were also genotyped on the Illumina Infinium HumanOmniExpressExome array. In parallel, we also leveraged a completely separate, non-overlapping cohort consisting of RNA-sequencing data from the post-mortem DLPFC tissue (SCZ=100, Controls=217) collected and generated on behalf of National Institute of Mental Health (NIMH) Human Brain Collection Core (HBCC). All data are downloadable through the CMC synapse.org panel. No statistical methods were used to pre-determine sample sizes but our sample sizes are the largest to be reported.

Data exclusions
The following exclusion criteria were pre-established: Individual sample level data were excluded on the basis of not detecting >80% editing sites shared with all samples. Brain tissue was not collected if cases had neuropathology related to Alzheimer’s disease and/or Parkinson’s disease, acute neurological insults (anoxia, strokes, and/or traumatic brain injury) immediately prior to death, or were on ventilators near the time of death.

Replication
Two brain regions were examined, including the DLPFC (Brodmann areas 9 and 46) and ACC. Results from this discovery cohort were then reproduced in a separate, non-overlapping DLPFC cohort generated through the National Institute of Mental Health (NIMH) Human Brain Collection Core (HBCC). For sample size, see above. Genome-wide concordance of altered RNA editing sites between discovery and validation cohorts was overall high. Some results did not validate in this separate cohort, including differences in ADAR2 expression levels; differential RNA editing analysis produced fewer FDR significant sites; some genes show enrichment for differentially edited sites while others do not; motif RBP enrichment analysis was less informative. We suspect these discrepancies are due to power issues in that our validation cohort is composed of two-fold fewer samples.

Randomization
The percentage of editing variance for each site was partitioned into the variance attributable to each variable using a linear mixed model implemented in the R package variancePartition version 1.12.08. Under this framework, categorical variables (i.e., sample site, biological sex) are modeled as random effects and continuous variables (i.e., individual age, PMI) are modeled as fixed effects. Each site was considered separately and the results for all sites were aggregated afterwards. This approach enabled us to rationally include leading covariates into our downstream analysis, which may ultimately have an influence on differential RNA editing analysis.

Blinding
Blinding was not possible as data analysis required sufficient insight into which samples corresponded to which groups in order to make rationale statistical comparisons.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.
Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | Antibodies            |
| ☒   | Eukaryotic cell lines |
| ☒   | Palaeontology         |
| ☒   | Animals and other organisms |
| ☒   | Human research participants |
| ☒   | Clinical data         |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | ChIP-seq              |
| ☒   | Flow cytometry        |
| ☒   | MRI-based neuroimaging |

Eukaryotic cell lines

Policy information about cell lines

| Cell line source(s)               | State the source of each cell line used. |
|-----------------------------------|----------------------------------------|
| Authentication                    | Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated. |
| Mycoplasma contamination          | Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination. |
| Commonly misidentified lines      | Name any commonly misidentified cell lines used in the study and provide a rationale for their use. |

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| Laboratory animals               | For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals. |
| Wild animals                     | Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals. |
| Field-collected samples          | For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field. |
| Ethics oversight                 | Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

| Population characteristics | From the CommonMind, a total of 470 ACC (SCZ=225, Control=245) and 540 DLPFC (SCZ=254, Control=286) RNA-sequencing samples were analyzed. These sample comprised a total of 358 unique SCZ cases and 380 unique controls (occasionally, both brain regions were dissected from the same individual whereas other individuals will have dissections from only one region). Control samples are fairly well matched with SCZ cases based on sample site, gender, ethnicity age and brain weight. Other covariates, including PMI, RIN and pH varied slightly between SCZ cases and controls. From the NIMH HBCC, a total of 317 DLPFC samples (SCZ=100, Control=217) were used as validation. These samples were well matched for gender, ethnicity and brain weight, but slight differences were observed among age, PMI, RIN and pH between SCZ cases and controls. These factors were all adjusted in our linear models to aid in controlling for their possible confounding effects on RNA editing. All of this information can be found in Supplemental Figure 1. |
| Recruitment                    | Postmortem tissue from schizophrenia (SCZ) and bipolar or other affective/mood disorder (AFF) cases were included if they met the appropriate diagnostic DSM-IV criteria, as determined in consensus conferences after review of medical records, direct clinical assessments, and interviews of family members or care providers. Cases were excluded if they had neuropathology related to Alzheimer’s disease and/or Parkinson’s disease, acute neurological insults (anoxia, strokes, and/or traumatic brain injury) immediately prior to death, or were on ventilators near the time of death. |
| Ethics oversight               | The Icahn School of Medicine at Mount Sinai, the University of Pennsylvania, and the University of Pittsburgh. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.