UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Analysis of RNA Editing in Brain Diseases

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Bioinformatics

by

Mudra Choudhury

2022

ABSTRACT OF THE DISSERTATION

Computational Analysis of RNA Editing in Brain Diseases

by

Mudra Choudhury

Doctor of Philosophy in Bioinformatics

University of California, Los Angeles, 2022

Professor Xinshu Xiao

It is well-established that RNA editing plays significant roles in brain function. Recent studies uncovered many RNA editing events with aberrant editing levels in neuropsychiatric and neurological diseases, such as Schizophrenia (SCZ) and Alzheimer's disease (AD). Yet, the underlying mechanisms of dysregulation of these RNA editing abnormalities and their contribution to disease processes are generally unknown. In this dissertation, we carried out in-depth studies of RNA editing to gain an improved understanding of its implications in brain diseases.

To better understand the roles of RNA editing in SCZ, we conducted global *de novo* RNA detection to probe editing differences between disease and control subjects in four independent SCZ cohorts. We observed reproducible and significantly reduced editing levels (i.e., hypoediting) in SCZ and an enrichment of dysregulated sites in genes involved in mitochondrial functions. Furthermore, we carried out experimental studies to

characterize the functional roles of dysregulated RNA editing located in coding and non-coding regions. These studies again highlighted the important contributions of RNA editing to mitochondrial function.

Next, we examined the relationship between genomic variation, RNA editing and other post-transcriptional processes in SCZ via quantitative trait loci (QTL) analyses. Detection of editing QTL (edQTL), splicing QTL (sQTL), and expression QTL (eQTL) in the CommonMind SCZ cohort revealed both common and distinct loci among the three types of QTL. In addition, we investigated each QTL-type in both European (EU) and African American (AA) populations, and observed that AA-specific QTL were associated with larger effect sizes. Finally, we demonstrated the disease relevance of QTL through their colocalization with the GWAS summary statistics of SCZ, bipolar disorder (BPD), and major depressive disorder (MDD), respectively.

Recently, RNA editing was reported to significantly impact the immunogenicity of double-stranded RNAs (dsRNAs). Motivated by this relationship, we sought to examine dsRNA expression and RNA editing dysregulation in disease. To this end, we implemented a bioinformatic pipeline to predict dsRNA regions transcriptome-wide. Using RNA-seq data of AD patients and controls, we identified global upregulation of dsRNAs and downregulation of RNA editing in AD, a disorder for which emerging evidence supports the importance of inflammation and innate immunity in disease mechanisms. Interestingly, while differentially expressed dsRNAs and reduced RNA editing are observed in nonoverlapping loci, they both significantly associated with interferon (IFN) response. Our data suggest that reduced RNA editing and increased

dsRNA expression collectively contribute to increased IFN response in AD, although through independent transcripts.

The dissertation of Mudra Choudhury is approved.

Qing Zhou

Roel Ophoff

Michael J. Gandal

Xinshu Xiao, Committee Chair

University of California, Los Angeles

2022

To my loved ones.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SUPPLEMENTARY FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY TABLES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Xinshu Xiao, for being an amazing role model and passionate researcher. Her hands-on approach to mentorship shows her dedication to her students, and ensures that her mentees develop into capable scientists. Despite having many responsibilities, she always makes time to meet or provide the appropriate help. I am especially inspired by her hard work and detail-oriented nature. Her support throughout my PhD has been crucial to my success as a student.

I would also like to thank the members of my committee, Dr. Michael Gandal, Dr. Roel Ophoff, and Dr. Qing Zhou. I appreciate their willingness to meet and all their insightful comments. Many times, each committee member has provided me with helpful suggestions that I incorporated into my projects. Their questions and advise during our meetings have strengthened my work.

Next, I would like to thank all current and former members of the Xiao Lab. I very much appreciate the collaborative environment of our lab. I have always felt that the lab is a safe space and that my colleagues are my friends. Thank you all so much for taking time to have lunch together and help me in my day-to-day research. I have learned so much from you all. I especially want to thank Stephen Tran, who was my student mentor when I joined the lab. Stephen jump-started my project in RNA editing in schizophrenia, and was a pillar of support for the first few years of my PhD.

Finally, I would like to thank all the family and friends who have helped me throughout my time at UCLA. First, to my mom, dad, and brother, who provided constant emotional support through the years. Thank you to all the friends I made at UCLA, who could relate to me during my struggles and helped me feel less alone. Thank you to the

friends I made outside of UCLA, who took my mind off of work. Although there are too many names to list, you all contributed more to my journey than you know. Thank you to Seetal and Kasturi Ai, who opened their homes to me when I had to commute to UCLA from San Diego after the pandemic. Your big hearts were a huge part of my PhD completion. Finally, a very big thank you to my partner of 10 years, Suchit. I cannot tell you how much I appreciate all that you have done to support me over the years. I am grateful to have you in my life. You make me better in my career, and more importantly, as a person.

This work was partially supported by NIH-NCI T32LM012424 (Biomedical Big Data Training Grant at UCLA, 2018-2020).

Chapter 2 is a version of a manuscript submitted for publication: Choudhury M, Fu T, Amoah K, Jun HI, Chan TW, Park S, Walker D, Bahn JH, Xiao X. Widespread RNA hypoediting in schizophrenia and its relevance to mitochondrial function. M.C, J.H.B. and X.X. designed the study with inputs from all other authors. M.C., T.W.C, and K.A. conducted bioinformatic analyses. H.I.J., J.H.B, and T.F. conducted molecular and cellular experiments. S.P. and D.W. contributed to experimental methods. All authors contributed to the writing of the paper. X.X. was the principal investigator. I would like to acknowledge Stephen Tran, who provided support in the bioinformatics methods.

Chapter 3 is currently being prepared as a manuscript for submission: Choudhury M, Xiao X. The genetic landscape of RNA editing and other aspects of gene expression in schizophrenia. M.C. and X.X. designed the study and wrote the paper. M.C. performed all analyses. X.X. was the principal investigator. I would like to acknowledge Dr. Michael Gandal, who had many helpful suggestions for QTL detection.

Chapter 4 is currently being prepared as a manuscript for submission: Choudhury M, Huang E, Xiao X. Genome-wide alteration of dsRNA expression and RNA editing in Alzheimer's disease. All authors contributed to study design and writing the manuscript. M.C. developed the dsRNA detection pipeline and conducted de novo differential RNA editing analysis. M.C. and E.H. identified differential dsRNAs in AD. E.H. conducted the association between AEI and AD severity. X.X. was the principal investigator.

VITA

Education and Employment

2012 – 2016	B.S. in Bioresource Research (Bioinformatics/Genomics option), Oregon State University Honors College, Corvallis, OR
2014 – 2016	Research Assistant, Department of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR
2017 – 2022	Graduate Student Researcher, Bioinformatics IDP, UCLA, Los Angeles, CA
Summer 2020	Computational Research Summer Intern, Bristol Myers Squib San Diego, CA
Fall 2020	Teaching Assistant, Department of Microbiology, Immunology and Molecular Genetics, UCLA, Los Angeles, CA

Honors and Awards

2018 - 2020	NIH Biomedical Big Data (BBD) Training Grant
June 2019	UCLA/NIH Biomedical Big Data Training Grant Data Challenge (1st)
May 2019	UCLA GATP-BWF-BBD Joint Poster Symposium (1st place poster)
2016 - 2018	UCLA Graduate Dean's Scholar Award
April 2017	UCLA Bioinformatics Retreat, Best Abstract
2012 - 2016	Oregon State University (OSU) Honors Promise Scholarship
2012 - 2016	OSU Academic Achievement Award

Winter 2015	OSU Undergraduate Research, Innovation, Scholarship and Creativity Grant
Winter 2015	OSU DeLoach Work Scholarship

Publications

Choudhury M, Fu T, Amoah K, Jun HI, Chan TW, Park S, Walker D, Bahn JH, Xiao X. *Widespread RNA hypoediting in schizophrenia and its relevance to mitochondrial function*. (Submitted manuscript under review)

Choudhury M, Xiao X. *The genetic landscape of RNA editing and other aspects of gene expression in schizophrenia.* (Manuscript in preparation)

Choudhury M, Huang E, Xiao X. *Genome-wide alteration of dsRNA expression and RNA editing in Alzheimer's disease.* (Manuscript in preparation)

Khan A., Bagley J., LaPierre N., Gonzalez-Figueroa C., Spencer T., **Choudhury M**., Xiao X. Eskin E., Jentsch J., Smith D. *Genetic pathways for longitudinal cocaine self-administration in mice.* (Manuscript in preparation)

Chan TW, Dodson J, **Choudhury M,** Xiao X. *Candidate RNA editing events relevant to immunotherapy response in melanoma.* (Manuscript in preparation)

Liu Z, Quinones-Valdez G, Fu T, **Choudhury M**, Reese F, Mortazavi A, Xiao X. *L-GIREMI uncovers RNA editing sites in long-read RNA-seq.* BioRxiv, doi:https://doi.org/10.1101/2022.03.23.485515; Preprint. (Submitted manuscript under review)

Choudhury, M. & Ramsey, S. A. *Identifying Cell Type-Specific Transcription Factors by Integrating ChIP-seq and eQTL Data-Application to Monocyte Gene Regulation.* Gene Regul. Syst. Bio. 10, 105–110 (2016).

CHAPTER 1 - Introduction

1.1 RNA EDITING OVERVIEW

1.1.1 Prevalence of RNA Editing

RNA editing, the alteration of targeted nucleotides on RNA molecules, notably reshapes our understanding of the central dogma of biology, which explains the transfer of information from DNA to RNA to protein. Most RNA editing sites involve the nucleotide conversion of adenosine to inosine (termed A-to-I editing) on double stranded pre-mRNA and are catalyzed by adenosine deaminase (ADAR) enzymes[1]. As the inosine nucleotide is interpreted as guanosine by cellular machinery, A-to-I editing is synonymously termed A-to-G editing. An alternative and less prevalent form of editing involves the substitution of cytosine with uracil (C-to-U), which is conducted by the apolipoprotein B mRNA editing catalytic polypeptide-like enzymes (APOBECs)[2].

Endogenous modification of RNA nucleotides has been observed across species, from single-celled protozoa to humans[3,4]. Furthermore, ADAR enzymes responsible for A-to-I editing originate in early metazoan evolution[3,5]. These proteins are highly conserved across a wide range of evolutionary scale, with considerable similarities among *Drosophila*, various cephalopods, mice, chimpanzees, and humans[4]. Nevertheless, humans exhibit a greater prevalence of RNA editing relative to other species[5].

In addition to its preservation across species, RNA editing is pervasive throughout the human genome. While the majority of RNA editing occurs in introns and takes place

in the nuclear fraction of the cell, edited loci are observed in various regions of the transcriptome and in alternative subcellular locations[6]. ADAR binding occurs predominantly in repetitive elements, such as Alu regions that occupy 10% of the human genome[7,8]. Consequently, RNA editing is estimated to occur in the majority of human genes at over 100 million loci, most of which are in noncoding regions with unknown functions[9]. The ubiquity of RNA editing across the genome and its conservation among species indicate the significance of its role in post-transcriptional regulation.

1.1.1 Functional Consequences of RNA Editing

RNA modifications confer diversity to their transcripts, affecting splicing, stability, transcript localization, gene expression, and downstream protein function[10–14]. For example, RNA editing in 3' splice site regions of highly conserved exons can cause dramatic reductions in splice site strength[10]. In addition, steady state transcript levels of disease-associated genes can be regulated through miRNA targeting of edited or unedited transcripts[15]. Finally, RNA editing can cause nonsynonymous changes to protein coding regions, such as the serine to glycine amino acid change in the brain-specific splicing factor, Nova1, which regulates ~700 alternatively spliced exons in genes involved with synaptic function and axon guidance. Although the edited locus does not affect the protein's splicing activity, it affects stability by doubling the protein's half-life[13]. These examples illustrate the importance of RNA editing in diversifying transcriptome complexity and regulation.

Furthermore, editing through ADAR proteins is known to be involved in the innate immune response in many species[16]. The ADAR gene in mammals gives rise to two

protein isoforms: a p110 isoform that is ubiquitously expressed, and a larger p150 isoform that is interferon-inducible[17]. Interestingly, Adar p150$^{-/-}$ mice are embryonically lethal due to their inability to protect against viral infection[18], similar to the phenotype observed in Adar$^{-/-}$ mice that lack both the p110 and p150 isoforms. Moreover, the Adar p150 protein is predicted to regulate the interferon (IFN) response cascade[18,19], and can edit viral dsRNAs to prevent sustained IFN expression and apoptosis during infection[20]. Overall, ADAR expression and RNA editing show multi-faceted roles in immune response.

1.1.2 The role of dsRNA editing in interferon response

Editing occurs in double-stranded RNA (dsRNA) regions that possess structural similarity to exogenous viral RNAs[16,21]. The RNA editing mechanism is utilized by the cell to distinguish between endogenous and exogenous RNAs, the latter of which can activate immune response[21]. The conversion of adenosine to inosine replaces the stable A:U base pair with the less stable I:U, introducing structural irregularities to the modified RNA duplex. Such structural changes prevent their recognition by viral RNA sensors, such as the Melanoma Differentiation-Associated factor 5 (MDA5), whose interaction with dsRNAs elicits antiviral signaling. Thus, RNA editing effectively marks endogenous dsRNA as “self” RNA and prevents unwanted IFN response[21].

Aberrant ADAR expression and editing can affect many cellular processes, such as dsRNA accumulation and innate immunity[22,23]. Hence, it is especially important to investigate the consequences of dysregulated RNA editing and dsRNA expression in disorders where immune response plays a role[24]. Indeed, decreased RNA editing and increased load of endogenous dsRNAs have been reported in the autoimmune disorder

psoriasis[25], although the mechanisms underlying such changes and their functional implications in the disease remain unknown.

1.2 DYSREGULATED RNA EDITING IN THE BRAIN

Prior to the availability of next-generation sequencing, RNA editing studies were focused on a small number of editing sites with significant roles in neuronal function[26]. The first discovered A-to-G editing site was an RNA-DNA difference (RDD) causing a glutamine (Q) to arginine (R) amino acid change in the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor subunit GluA2 in the brain[27]. This Q/R change (CAG-to-CGG) rendered the AMPA receptor impermeable to the Ca^{2+} ion[27]. Due to its interesting functionality, this site has been the focus of many in-depth studies since its discovery[28–31]. The Q/R GluA2 locus is highly edited (> 90%) in neurons in the adult CNS[29], and GluA2 subunits are almost all completely edited in embryonic and postnatal brains[30,31]. In mice, a reduction of the Q/R editing level to 75% leads to epilepsy, seizures, and premature lethality[32–35]. Moreover, decreased editing in GluA2 was found in motor neurons of patients with Amyotrophic Lateral Sclerosis (ALS) and is believed to impact motor neuron function[36]. Since the discovery of GluA2 editing, aberrant RNA editing in similar glutamate receptors (e.g. GRIK1 and GRIK2)[37], serotonin proteins (e.g. 5-HT(2C) receptor)[38], and other loci has been uncovered in multiple human neuropsychiatric diseases, such as Schizophrenia (SCZ), Autism Spectrum Disorder (ASD), Bipolar Disorder (BPD), and Alzheimer's Disease (AD)[38–42].

Following the advent of high-throughput sequencing technologies, the landscape of RNA editing has been greatly expanded to encompass diverse diseases, tissues and

cell types[43–45]. RNA editing is a cell-type specific mechanism that exhibits the highest prevalence in the brain relative to other tissues[46], motivating its investigation in brain-related disease. While earlier studies have exemplified the importance of RNA editing in neurological disorders, many investigated only a few sites or did not characterize the functional consequences of differential editing between patient and control samples. Complete characterization of RNA editing differences has not been achieved for most neurological and neuropsychiatric diseases. Moreover, many edited loci remain undiscovered and their functional roles unexplained. These deficiencies highlight the need for further studies of RNA editing sites in brain diseases on the global and local scales.

1.3 RNA EDITING IN SCZ

1.3.1 Dysregulated RNA editing in SCZ

SCZ is a neuropsychiatric disorder characterized by delusional thinking, hallucinations, anxiety, and paranoia. Both genetic and environmental causes, such as stress, substance use, and maternal perinatal infection, can trigger the illness[47]. As SCZ is highly heritable (having a heritability score of 0.8)[48], many studies have aimed to identify its genetic basis and the biological pathways implicated in its progression[47,49,50]. Genome-wide association studies (GWAS) have highlighted genetic variants that appear to collectively influence the disorder[49,50]. However, the underlying biological mechanisms, especially those beyond genetic mutations involved in shaping the disorder, are largely unknown. Consequently, the investigation of SCZ at the transcriptional and post-

transcriptional levels is increasing in popularity, which may yield novel insights into disease etiology[51,52].

One such post-transcriptional mechanism is RNA editing, which has been found to be dysregulated in SCZ and related neurological disorders, such as BPD[38,39,53]. For example, an editing site in the glutamate receptor, ionotropic kainate 2 (GRIK2) gene, causes a nonsynonymous I/V change with significant downstream functional consequences in the brain[39]. Similar to the GluA2 Q/R site, this locus modulates cellular Ca^{2+} permeability, and its dysregulation is predicted to contribute to the increased intracellular levels of Ca^{2+} observed in patients with BPD[39]. In addition to a handful of known functional editing sites, a global study of RNA editing in SCZ has examined its underlying genetic basis through RNA editing quantitative trait loci (edQTL) detection, and highlighted the disease relevance of such loci[53].

While these studies demonstrate the importance of RNA editing in neuropsychiatric disorders, the potential contribution of RNA editing to disease-related mechanisms is largely unknown. Additionally, investigation and comparison of global RNA editing in different ancestral groups has not been conducted. To address these gaps, Chapter 2 of this work methodically identifies *de novo* RNA editing sites in Europeans (EU) and African Americans (AA) in multiple SCZ cohorts, and characterizes the molecular pathways affected by differentially edited loci. We observed a reproducible global hypo-editing trend in SCZ for EU individuals. Importantly, we uncovered the potential connections of many aberrant RNA editing sites to mitochondrial function, which is known to be dysregulated in the disease[54]. We validated recoding RNA editing sites in

the mitochondrial gene, MFN1, and showed their impact on mitochondrial morphology and function.

1.3.2 Genetic basis of RNA editing in SCZ

Since SCZ is predicted to have a strong genetic basis[48], many studies focused on the relationship between genetic variation and gene expression, splicing, or other cellular processes in the disease[53,55,56]. Nonetheless, as an epitranscriptomic process that rapidly responds to environmental stimulus[57] and affects brain-related pathways, RNA editing may have close relevance to the mechanisms underlying SCZ. However, our understanding of the genetic basis for RNA editing in SCZ remains limited[53]. Previous analysis of quantitative trait loci associated with RNA editing (edQTL) by Breen et. al. reported a number of genetic variants associated with RNA editing loci, of which 11 colocalized to SCZ GWAS loci[53]. This study motivates further characterization of edQTL in SCZ, including such investigations in multiple ethnicities.

In Chapter 3, we aimed to further elucidate the association between RNA editing and genetic variation in SCZ for individuals of both EU and AA descent. In addition to edQTL detection, we conducted eQTL and sQTL analyses of the same cohort to compare the resulting loci and their targets. All QTL types showed higher effect sizes in AA-specific loci. Furthermore, we investigated RNA binding proteins (RBPs) that may regulate RNA editing or RNA splicing. We also conducted colocalization analysis of the QTL to GWAS loci from multiple brain disorders. While the three types of QTLs shared significant overlap, they were largely distinct and showed relevance to SCZ, major depressive disorder (MDD), and BPD.

1.4 DSRNA AND RNA EDITING DYSREGULATION IN ALZHEIMER'S DISEASE

RNA editing dysregulation in dsRNA regions is prevalent in a wide range of human diseases in various stages of human lifespan[58,59]. The functional relevance of RNA editing in brain development[60] makes it a significant focal topic for neurodevelopmental disorders including SCZ. However, its additional relevance to aging[57,61], environmental stress[57], and innate immunity[62] makes it equally interesting for neurodegenerative diseases, such as Alzheimer's disease (AD).

1.4.1 Characterization of Alzheimer's disease

AD is a common neurodegenerative disorder that is currently the seventh leading cause of death, with the number of cases almost doubling every five years[63]. It is characterized by initial memory impairment and cognitive decline, which can affect behavior, speech and the motor system[64]. Having a mean onset-age of 65 years-old[63], AD is one of the most common causes of dementia[64]. The disease's defining neuropathological characteristics include the accumulation of extracellular amyloid-beta ($A\beta$) plaques and intracellular neurofibrillary tangles (NFTs), loss of synapses, and neuronal death[65]. A definitive diagnosis of AD often requires examination of cerebrospinal fluid (CSF), positron emission tomography (PET) biomarkers, and newer criteria that aid diagnosis in living patients[66,67]. Due to its complex neuropathology, various metrics have been developed to quantify AD severity[68]. Through neuropsychological tests, an individual may be assigned a cognitive index[69]. Other often-used AD severity metrics include the CERAD score as a quantitative measure of the brain's $A\beta$ neuritic plaques[70], and Braak stages reflecting the amount of NFTs[71].

In addition to accumulation of plaques and tangles, AD is associated with increased neuroinflammation[72]. Persistent immune response in the brain has been shown to facilitate and exacerbate disease progression[72]. For example, the accumulation of Aβ plaques can activate microglia expressing IFN stimulated genes (ISGs)[72]. Furthermore, the inflammatory response evoked by Aβ plaques can consequently expedite disease progression through the onset of tau pathology[72]. In contrast, IFN blockade in mouse models of AD reduces microglia activation and rescues synapse loss in vivo[72,73]. Overall, IFN dependent processes are integral to the neuropathology of AD. While IFN is primarily activated through innate immune sensors recognizing viral and self dsRNAs[74], the direct mechanisms by which IFN is activated in AD remains unclear.

1.4.2 RNA Editing and dsRNA expression in AD

Dysregulated RNA editing in dsRNA regions has been shown to converge on brain and immune related pathways in AD[75,76]. While previous studies have provided valuable insights for RNA editing aberrations in the disease, the functional consequence of dysregulated editing in AD is largely uncharacterized. Furthermore, the relationship between editing dysregulation, dsRNA expression, and immune response in AD is unknown. The work presented in Chapter 4 aims to address these knowledge gaps. We detected editing enriched dsRNAs differentially expressed between AD and control brains and examined dysregulated editing in AD. We demonstrated that both of these processes, while occurring at distinct loci, associate with ISG expression.

1.5 SUMMARY

Previous literature has demonstrated that dysregulated RNA editing is an important molecular signature in both neurodevelopmental and neurodegenerative diseases. However, many questions remain regarding the functional consequences of RNA editing and its impact on disease pathology. In this dissertation, we aimed to better characterize RNA editing in brain disorders of interest, specifically SCZ and AD.

CHAPTER 2 - Widespread RNA hypoediting in schizophrenia and its relevance to mitochondrial function

2.1 ABSTRACT

RNA editing, the endogenous modification of nucleic acids, is altered in genes with significant neurological function in schizophrenia (SCZ). While RNA editing sites have been characterized in neuronal genes, such as glutamate and serotonin receptors, its role in SCZ-associated mitochondrial dysregulation has not been explored. Moreover, the global profile and molecular functions of disease-associated RNA editing remain unclear. Here, we analyzed RNA editing in postmortem brains of four SCZ cohorts and uncovered more than 26,000 unique differential editing sites between SCZ and controls. We discover a significant hypoediting trend in patients of European descent and report a set of SCZ-associated editing sites via WGCNA analysis shared across cohorts. We show that differential editing patterns reflect editing differences in multiple cell types, including neurons, rather than cell compositional differences between conditions. Robustly identified differential editing sites were enriched in functionalities related to the brain and mitochondria (i.e. glutamate receptor activity and mitochondrial fragmentation in apoptotic processes). Indeed, using massively parallel reporter assays and bioinformatic analyses, we see differential 3' UTR editing sites affecting host gene expression were also enriched for mitochondrial functions, such as cellular respiration. Furthermore, we characterized the impact of two recoding sites in the mitofusin 1 (*MFN1*) gene and showed their functional relevance to mitochondrial fusion, cytochrome C release, and cellular

apoptosis. To our knowledge, this is the first study to show replicable global reduced editing in SCZ and investigate its crucial molecular effects on mitochondrial activity.

2.2 INTRODUCTION

RNA editing, the alteration of targeted nucleotides on RNA molecules, notably reshapes our understanding of the central dogma of biology. Most RNA editing sites involve the nucleotide conversion from adenosine to inosine (termed A-to-I editing) on double-stranded pre-mRNAs catalyzed by the adenosine deaminase (ADAR) enzymes[1]. As inosines are interpreted as guanosines by the subsequent cellular machineries, A-to-I editing is synonymously termed A-to-G editing. An alternative but less prevalent form of editing involves the substitution of cytosine with uracil (C-to-U or C-to-T editing), which is conducted by the apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) enzymes[2]. Occurring in both coding and non-coding regions, RNA editing greatly diversifies the transcriptome[6]. In addition to altering protein sequences, RNA editing may affect many other processes, such as splicing[10], RNA stability[77], and translation[9]. Thus, identifying functionally meaningful editing sites can help to understand their primary biological roles and, moreover, elucidate how dysregulated editing contributes to various disorders.

Previous studies of RNA editing have shown the significance of aberrant editing in neurological diseases[78]. One such disease is Schizophrenia (SCZ), in which differences in RNA editing have been profiled in both non-coding and coding regions[53]. SCZ is a neuropsychiatric disorder characterized by delusional thinking, hallucinations, anxiety, paranoia, and a variety of other psychiatric symptoms[79]. The disorder is influenced by

both genetic and environmental factors, such as stress, substance use, and maternal perinatal infection[47]. As SCZ is highly heritable, many studies have aimed to identify the genetic basis of the disease and understand the biological pathways implicated in its progression[47]. Genome wide association studies (GWAS) have identified genetic variants that may collectively influence the disorder[49,80]. However, the underlying biological mechanisms involved in shaping the condition beyond genetic mutations are largely unknown. Consequently, investigation of SCZ at the transcriptional and post-transcriptional level has gained traction, as this may yield novel insights into disease etiology[51,52].

Various transcriptomic regions with altered RNA editing have been catalogued in SCZ and many of its related brain disorders. For example, an I/V editing site in the glutamate receptor, ionotropic kainate 2 (GRIK2), is known to cause a nonsynonymous change in its protein. GRIK2 modulates cellular Ca^{2+} permeability, and its dysregulation contributes to increased intracellular Ca^{2+} levels observed in patients with bipolar disorder, a mental disease sharing significant neuropathology with SCZ[81,82]. Additionally, global profiling of RNA editing in SCZ has shown dysregulation of hundreds of RNA editing sites, including those in genes involved in translation initiation and AMPA glutamate and ionotropic receptors[53]. These studies confirm the importance of editing in this neuropsychiatric disorder and motivate further in-depth investigations on the global profile, regulation and function of RNA editing in SCZ and related disorders.

In this study, we characterized the RNA editomes in brain samples of SCZ patients and controls in multiple cohorts from the PsychENCODE consortium via *de novo* detection of RNA-DNA differences. Comparing data from SCZ and control individuals, we

identified 26,841 unique differential editing sites. Importantly, we observed a significant trend of hypoediting in SCZ, which was reproduced in 3 of the 4 cohorts of European individuals. Moreover, our study uncovered close relevance of RNA editing dysregulation to mitochondrial function from two perspectives: RNA editing in 3' UTRs of mitochondria-related genes and the impact of two recoding editing sites in the mitofusin 1 (MFN1) protein. Together, our study delineates a comprehensive landscape of RNA editing in SCZ, reports a replicable hypoediting bias in SCZ for the first time, and reveals novel functional relevance of RNA editing in mitochondria-related processes.

2.3 RESULTS

2.3.1 RNA Editing sites detected in SCZ patients and controls of the BRAINGVEX data

We first aimed to identify global RNA editing profiles in SCZ to further understand novel functional pathways dysregulated in the disease. We obtained brain frontal cortex (FC) RNA-seq data of 170 samples from the BrainGVEX cohort in the PsychENCODE consortium[83] (Supplementary Table 2-1). The RNA-seq libraries were generated using rRNA depleted total RNA[83]. On average, 41 million reads were uniquely mapped per sample (Supplementary Fig. 2-1a). Following sequence alignment and quality control procedures, 65 SCZ and 67 control samples were retained (Methods), the vast majority (96%) being from European populations. We ensured that metadata variables and data-related metrics (e.g., age, gender, RIN, PMI and sequencing depth) did not significantly differ between SCZ and control individuals (Supplementary Fig. 2-2a, b).

RNA editing sites were detected using our previously developed *de novo* RNA editing detection pipeline[84–86]. As editing sites are often close to one another within a

region of the transcriptome, we further implemented a method to detect sites within these "hot-spots" that may be otherwise missed by conventional mapping software[41,87] (Methods). In total, we detected 4,576,706 RNA editing sites in the FC of the BrainGVEX cohort. In order to avoid sites with rare occurrences, we discarded those with nonzero editing in less than 10% of the samples. Following this filter, a total of 255,812 sites were retained, referred to as "common" RNA editing sites (Fig. 2-1a). Of these common sites, 212,532 (83%) overlapped with known RNA editing sites cataloged in the REDIportal database[88] (Supplementary Fig. 2-1b).

Almost all samples had greater than 95% of A-to-G sites out of all common RNA editing sites detected in the respective sample, demonstrating the high accuracy of the *de novo* detection pipeline (Fig. 2-1b, Supplementary Fig. 2-1c). Consistent with previous studies[86,89], the majority of editing sites resided in ALU elements (Supplementary Fig. 2-1d) and intronic regions (Supplementary Fig. 2-1e). As expected, the number of sites detected per sample correlated approximately with the total read coverage of each sample (Supplementary Fig. 2-1f). We calculated common site RNA editing average (CREA) per sample using all common sites with at least 5 reads. An ALU specific CREA was also calculated using common sites in ALU regions. Both CREA and ALU-CREA had significantly positive correlations with ADAR2 expression and negative correlations with ADAR3 expression, while both insignificantly correlated with the expression of ADAR1 (Supplementary Fig. 2-1g).

2.3.2 Differential editing analysis in BrainGVEX data reveals reduced editing in Europeans with SCZ

Using our previous methods[41], we detected 13,997 differential editing sites between SCZ and controls in the BrainGVEX cohort (Fig. 2-1a, Methods). To represent overall editing level of these differential editing sites per sample, we calculated the average editing ratio of all differential sites with ≥ 5 reads in each sample, hereby referred to as the differential site RNA editing average (DREA). In comparing SCZ and control samples, we observed significantly lower DREA values in the SCZ samples (Fig. 2-1c). The SCZ and control samples were clearly segregated when clustered based on the differential editing sites (Fig. 2-1d). In addition, differential sites that were hypoedited in SCZ greatly outnumbered those that have increased editing in the disease (Fig. 2-1e). Gene ontology (GO) analysis revealed that differential editing sites occurred in genes involved in various brain-related pathways such as motor *neuron axon guidance*, *synaptic transmission* and *ion transport* (Supplementary Fig. 2-1h).

Among the differential editing sites, a number of them cause nonsynonymous amino acid changes, stop loss, or stop gain (recoding sites) (Fig. 2-1f). Some of these sites are known to have relevance in neuropsychiatric diseases. One example is the aforementioned I/V recoding site in GRIK2, a gene encoding the well-known ionotropic glutamate receptor implicated in mood disorders. In SCZ, we observed hypoediting of this site (Fig. 2-1f), which is similar to that reported in bipolar disorders[81]. This hypoediting event can lead to an increase in intracellular Ca^{2+} levels[81]. Furthermore, we observed significantly decreased editing for sites that have not been studied previously in SCZ,

such as a C-to-T site in the mitochondrial fusion gene *MFN1* (Fig. 2-1f), which is further examined below.

2.3.3 RNA editing detection and differentially edited gene expression in single-nucleus RNA-seq

As RNA editing may be highly cell type-specific[90], we next asked whether the differential editing patterns in SCZ may have originated from specific cell types. To address this question, we analyzed single-nucleus RNA-seq data from prefrontal cortex of control samples in a previous study[91]. Six major brain cell types, excitatory and inhibitory neurons, astrocytes, oligodendrocytes, microglia, endothelial, and oligodendrocyte progenitor cells (OPC), were obtained from 24 samples. We examined gene expression and editing profiles of each cell type (by pooling cells of the same type). Excitatory and inhibitory neurons expressed the highest fraction of genes that were differentially edited in the bulk brain tissue of SCZ, followed by oligodendrocytes, astrocytes, and OPCs (Fig. 2-1g). Next, we asked whether the differentially edited sites in the bulk brain of SCZ were edited in each cell type. Among all observed editing sites in each cell type, the fraction of sites that were differentially edited in the bulk data is highest in neurons (Fig. 2-1h). These observations suggest that differential RNA editing observed in the bulk RNA-seq analysis may reflect editing differences in neurons to a larger extent than in other cell types.

Given the above RNA editing differences among cell types, we next asked whether cell type proportions were different between the bulk SCZ and control samples, which may contribute to the observed differential editing profiles. We used CIBERSORTx[92] to

calculate the relative proportion for five cell types of brain frontal cortex – neurons, oligodendrocytes, microglia, endothelial cells, and astrocytes – using gene expression of known cell type signature genes[93]. We observed an insignificant difference in cell type proportion between SCZ and control samples (Supplementary Fig. 2-1i). Thus, the differential editing observed between the two groups may not have arisen from cell type proportion differences. In contrast, using differential editing sites located in the signature genes of each cell type, we observed that the mean DREA was reduced in SCZ for the majority of cell types (Supplementary Fig. 2-1j). These observations further support the global hypoediting trend observed in the bulk data of SCZ.

2.3.4 Differential RNA editing analysis in additional cohorts

In order to validate global editing patterns discovered in the BrainGVEX cohort, we identified RNA editing sites in an additional cohort in the PsychENCODE database, referred to as the CommonMind Consortium (CMC)[94] (Supplementary Table 2-1). The same analysis methods used for the BrainGVEX data were adopted here. While the CMC cohort had a larger number of samples than the BrainGVEX cohort, close to half of the CMC subjects were above the age of 70 years old (an age group not present in the BrainGVEX cohort, Supplementary Fig 2-2a). Additionally, samples in the 70+ age group showed significantly lower RIN than the <70 age group, and their corresponding RNA-seq data had lower proportion of mapped mRNA bases and a higher proportion of intergenic bases (Supplementary Fig 2-3a-c). Considering the possibility of lower data quality from samples with 70+ age and focusing on the goal of validating the global editing patterns of BrainGVEX, we only used CMC samples from subjects below 70 years of age.

After quality control procedures and meta-data matching (Methods), 137 samples (79% European) were retained for RNA editing analysis and differential editing detection (Supplementary Fig. 2-4a, b). The total number of detected sites, common sites (edited in ≥10% of samples), differentially edited sites, and recoding sites are shown in Fig. 2-2a. Using the differential editing sites, we calculated the DREA for each sample, and observed a significant trend of reduced editing in SCZ (Fig. 2-2a), consistent with the observation in the BrainGVEX cohort.

Next, we analyzed data from two additional PsychENCODE cohorts, the CMC Human Brain Collection Core (HBCC) and the Lieber Institute for Brain Development (LIBD). Since these two cohorts included samples from European origin along with a sizeable number of samples of African American descent, we analyzed data from the two ethnicities separately. After quality control procedures, we analyzed 18 SCZ, 23 controls and 17 SCZ, 20 controls of European-descent from HBCC and LIBD respectively. Despite the relatively small sample size, we observed a significant hypoediting trend in SCZ of the HBCC cohort (Fig. 2-2b). No significant difference was detected in the LIBD data (Fig. 2-2c). Thus, RNA editing levels were significantly reduced in SCZ in three of the four European-dominant cohorts.

Additionally, we analyzed data from 48 SCZ, 52 controls (HBCC) and 18 SCZ, 26 controls (LIBD) of African American-descent. Opposite to the reduced editing in SCZ observed in European samples, a significant increase in DREA of SCZ relative to controls was detected in both cohorts (Supplementary Fig. 2-5a). Using CIBERSORTx, we did not observe a significant difference in cell type proportion between SCZ and controls (Supplementary Fig. 2-5b). While the DREA of signature genes of each cell type was

higher in SCZ than controls for most cell types in both cohorts, it only reached significance in astrocytes (HBCC) and oligodendrocytes (LIBD) (Supplementary Fig. 2-5c). Thus, we hypothesize that the hyperediting trend in SCZ of African American individuals may be due to ethnicity-related differences in RNA editing, a topic that needs to be further investigated.

2.3.5 WGCNA of edited loci yields robust modules of editing sites associated with SCZ

We next examined whether differential editing sites were shared between the BrainGVEX and CMC data (the HBCC and LIBD cohorts were not included due to their relatively small numbers of European subjects). While 314 overlapping differentially edited sites were observed between the cohorts, the total overlap did not reach statistical significance (Supplementary Fig. 2-6a). However, genes harboring differential editing sites showed a significant overlap, suggesting a similarity in functional pathways being differentially edited (Supplementary Fig. 2-6b).

To further evaluate SCZ-relevant editing sites shared between the two cohorts, we carried out the weighted gene co-expression network analysis (WGCNA)[95] on common RNA editing sites (Methods) in each cohort. The goal of this analysis was to identify RNA editing modules (i.e., sites with correlated editing levels across samples) that are associated with disease condition. For each cohort, WGCNA yielded multiple modules (Supplementary Fig. 2-7a, b), the eigengenes of which were then correlated with the disease condition while considering confounding meta-data covariates (age, gender, RIN, etc., see Methods). We observed that the largest module – labeled as the “turquoise” module – significantly correlated with SCZ in BrainGVEX and CMC, respectively

(Supplementary Fig. 2-7c). Importantly, editing sites in turquoise modules of the two cohorts significantly overlapped with each other when considering editing sites testable in both cohorts (Fig. 2-2d, Supplementary Table 2-2).

The two cohorts also shared a significant number of genes harboring turquoise module editing sites, as expected (Fig. 2-2e). Gene ontology (GO) analysis of the shared genes uncovered many pathways related to brain functionality and neuronal signaling, such as *ionotropic glutamate receptor activity* and *synapse assembly* (Fig. 2-2f). Interestingly, one of the top GO terms identified is related to *mitochondrial fragmentation in apoptotic process* (Fig. 2-2f). Together, these data support the existence of reproducible differential editing profiles in SCZ, many located in genes with functional relevance to brain function.

2.3.6 Enrichment of potentially functional 3'UTR editing sites in mitochondria-related pathways

Given the large number of differential editing sites in SCZ, it is important to investigate their functional relevance. To this end, we first focused on sites in the 3' UTRs given the observed relative enrichment of differential sites in this region (Supplementary Fig. 2-6c). Since 3' UTRs are enriched with cis-regulatory elements, RNA editing in the 3' UTRs may affect post-transcriptional gene regulation, for example, mRNA abundance, as shown in previous studies[77,96]. To experimentally screen for functional 3' UTR editing sites in regulating mRNA abundance, we performed a massively parallel reporter assay (MPRA), similar to those adopted to discover functional 3' UTR SNPs[97] (Fig. 2-3a, Methods). Although the edited version of the RNA editing site was hard-coded into the

DNA of the reporters, this design represents a reasonable approximation to assay for the impact of a 3' UTR editing site on mRNA abundance. This is because such regulation most likely occurs in the cytoplasm, independent of the process of RNA editing.

In the MPRA, we included a total of 770 differential editing sites located in 3' UTRs from the BrainGVEX, CMC, HBCC, or LIBD data. Comparing the relative enrichment of the unedited and edited versions of a site in the plasmid DNA input and expressed mRNA, we identified 214 editing sites (28% of 770 testable sites, located in 160 genes, Supplementary Table 2-3) that resulted in significant reporter expression changes (FDR ≤ 0.1 and $|\ln(\text{Fold Change})| \geq 0.1$, Fig. 2-3b; Methods). Thus, a relatively large fraction of 3' UTR editing sites may regulate gene expression post-transcriptionally. Interestingly, the GO analysis of genes harboring significant sites identified in the MPRA revealed a number of pathways related to mitochondrial function or translational regulation (Fig. 2-3c).

As a complementary approach, we examined the correlation between the editing levels of 3' UTR differential editing sites (tested in the MPRA) and the expression levels of their host genes in the respective cohorts (Fig. 2-3d). We found that 74 out of 741 bioinformatically testable editing sites were significantly correlated with gene expression (Methods). These sites were located in 65 genes, 30 of which overlapped with the genes with significant MPRA results ($p = 1.7e\text{-}46$, hypergeometric test). Consistent with the MPRA results, genes containing the 74 sites were significantly enriched with mitochondria-related pathways, such as *respiratory electron transport chain*, *mitochondrial inner membrane*, and *mitochondrial matrix* (Fig. 2-3c).

To further examine the potential function of genes whose 3' UTR editing was associated with gene expression (3' UTR EdEx genes), we analyzed functional pathways enriched among the union of 195 such genes discovered experimentally or bioinformatically. Via the GeneMANIA database[98] and Cytoscape visualization tool[99], we created a network of previously curated associations between the 195 3' UTR EdEx genes, as well as 20 additional related genes. We observed that the 3' UTR EdEx genes are highly interconnected, largely through genetic interactions (Fig. 2-3e). Among all modules of the network, mitochondrial processes harbored the most genes with the most significant enrichment level (Fig. 2-3e). Together, these analyses support that differential editing in the 3' UTR may affect mitochondria-related genes by altering their gene expression in SCZ.

2.3.7 RNA editing of MFN1 leads to reduced mitochondrial fusion

In addition to editing sites in non-protein coding regions, many of the protein recoding sites identified as differential in SCZ have interesting and diverse functionalities. As examples, we focused on two recoding sites in the gene *MFN1* that encodes a mitochondrial membrane protein essential for mitochondrial fusion[100]. One of the sites included a C-to-T recoding event in MFN1 with differentially reduced editing levels in SCZ (Fig. 2-4a). Another A-to-G editing site is located in the adjacent codon to the C-to-T site, which was not differentially edited in SCZ. We confirmed the presence of both RNA editing sites in human brain samples (Fig. 2-4b). Further investigation revealed that both editing sites detected in our analyses are conserved across species such as Macaca mulatta[101], mouse[102], and zebrafish[103] (Supplementary Fig. 2-8a). In addition, we observed that the

editing levels of the MFN1 C-to-T and A-to-G loci are correlated with each other, as editing of the A-to-G site is more likely if the C-to-T editing is present (Supplementary Fig. 2-8b). In mouse N2a cells, we confirmed that the C-to-T editing site is regulated by Apobec2 (Fig. 2-4c, Supplementary Fig 2-8c), further supporting the validity of this edited locus.

Next, we examined the impact of the two editing sites on mitochondrial fusion, a well-established function of MFN1[104]. We first used mouse embryonic fibroblast (MEF) cells with Mfn1 and Mfn2 double knockout (dKO). Overexpression vectors of Mfn1 and its edited versions (I328V from A-to-G editing, S329L from C-to-T editing and I328V+S329L double editing) were generated and introduced into dKO MEF cells. Expression of the Mfn1 proteins were confirmed via Western blot (Fig. 2-4d). The mitochondria morphology was examined via mitoTracker (Fig. 2-4e). Consistent with previous literature[104,105], we observed a severe fragmented mitochondria phenotype in Mfn1/2 dKO cells (Fig. 2-4e). In contrast, rescuing by the wildtype Mfn1 in the dKO background showed 48% long tubular mitochondria, indicating that Mfn1 overexpression restored mitochondrial fusion. Expression of single-edited Mfn1 mutant (I328V or S329L) led to moderate mitochondria morphologies encompassing fragmented, short, and long tubular phenotypes. Strikingly, the double-edited mutant (I328V+S329) induced significantly less long tubular (16%) and more short tubular (49%) mitochondrial phenotypes (Fig. 2-4e). Thus, our data suggest that the two recoding sites in Mfn1 reduced the protein's function in mitochondrial fusion.

As an alternative strategy, we created mutant HEK293T cells that carried the edited bases in their genome by prime editing[106]. Specifically, cells with a single mutation (corresponding to the I328V or S329L site) or double mutations (I328V+S329L) were generated. Clones with heterozygous mutations were retained (Supplementary Fig. 2-8d).

We confirmed that there was no change in MFN1 and MFN2 expression between WT and mutant cells (Supplementary Fig. 2-8e). Note that wildtype (WT) HEK293T cells do not exhibit RNA editing in either editing site (data not shown). In WT HEK293T cells, we observed 63% long tubular and 14% fragmented mitochondria (Fig. 2-4f). In contrast, the double mutant cells had a significantly higher fraction of fragmented (51%) and lower fraction of long tubular (5%) mitochondria, whereas the single mutant cells demonstrated intermediate phenotypes between the WT and double mutant cells (Fig. 2-4f). Thus, the edited versions of MFN1 induced significant reduction of mitochondrial fusion. This observation is largely consistent with that observed in mouse cells, despite the differences in the two systems (e.g., presence of endogenous MFN2 in HEK293T cells vs. absence of MFN2 in MEF cells).

To further investigate the impact of MFN1 editing on mitochondrial fusion, we performed a stress-induced mitochondrial hyperfusion test[107]. We treated HEK293T cells with cycloheximide, a widely used inhibitor for translational elongation, as a stress to induce mitochondrial hyperfusion, similarly as in previous studies[108]. We observed that 52% of WT cells had hyperfused mitochondria, whereas only 16% of double mutant cells underwent hyperfusion (Fig. 2-4g). Thus, cells with double-edited MFN1 had impaired hyperfusion response given cycloheximide, indicating a possible defect in adaptive stress response of these cells.

2.3.8 RNA editing of MFN1 affects cellular apoptosis

Mitochondria are central players in cell apoptosis. It is known that mitochondrial fission and fusion processes are closely implicated in apoptosis and loss of MFN1/2 leads

to increased apoptotic sensitivity[109,110]. Thus, we evaluated the potential impact of RNA editing of MFN1 on apoptosis. To this end, we measured the level of cytochrome C (Cyt C) in the cytoplasmic and mitochondrial fractions of the above WT and mutant HEK293T cells. Release of Cyt C from the mitochondria into the cytosol is a known hallmark of apoptosis[111]. As expected, in the WT cells, Cyt C was primarily localized in the mitochondrial fraction. In contrast, mutant cells with edited versions of MFN1 showed increased Cyt C in the cytosol, with the double edited cells showing the highest levels of Cyt C release (Fig. 2-5a).

To corroborate the above observations, we measured the level of apoptosis via propidium iodide (PI) staining. As shown in Fig. 2-5b, very few apoptotic cells were detected in the WT cells, as expected. In contrast, the mutant cells carrying one or both edits showed a significantly higher level of apoptosis, with the double edited cells being the highest. This observation is consistent with the results obtained via the Cyt C release assay. Thus, both experiments suggest that the RNA editing sites in MFN1 indeed affect cellular apoptosis.

2.4 DISCUSSION

Our study yields the most comprehensive investigation of dysregulated RNA editing in SCZ to date. For the first time, we uncovered a common trend of hypoediting in individuals with SCZ across multiple cohorts of primarily European-descent. Furthermore, we investigated the possible functional consequences of differential editing sites in 3' UTRs using MPRA. This experiment, complemented by bioinformatic analysis, revealed hundreds of 3' UTR editing sites that may alter mRNA expression. Importantly, many of

these functional editing sites are located in genes with relevance to mitochondria function. We also uncovered novel functions of two recoding editing sites in *MFN1*, an important gene for mitochondrial fusion.

Previous studies of neurological and neuropsychiatric disorders demonstrated the relevance of RNA editing to brain functionality. For example, our group's work on autism spectrum disorders (ASD) revealed RNA editing changes in genes involved in glutamate receptor activity and synaptic transmission[41], which are analogous to those uncovered in this study. Importantly, a significant hypoediting trend was observed in ASD. Other studies also identified reduced editing in brain disease for specific sites, such as those in the glutamate receptor GRIK2 in bipolar disorder[39] and the AMPA/kainate receptor GluR2 in amyotrophic lateral sclerosis (ALS)[112]. However, a recent global analysis of RNA editing in SCZ did not observe this distinct hypoediting trend for their reported differential sites, although reduced editing in AMPA-type glutamate receptors and postsynaptic density proteins was observed[53]. A number of factors may have contributed to the distinction between ours and the previous study. We implemented strict quality control procedures, focused on a consistent age range for all cohorts (<70 years), and carried out a rigorous de novo detection of RNA editing sites that allowed us to identify a large number of loci. Furthermore, the European and African American cohorts showed opposite overall trend of editing changes in SCZ, which were not analyzed separately in the previous study[53].

In contrast to the hypoediting trend in SCZ of the European cohorts, the African American cohorts showed an overall hyperediting bias in SCZ. Our analysis suggests that the hyperediting trend in African American samples exists in multiple cell types, as shown

when only cell type-specific marker genes were analyzed. These results suggest the possible existence of mechanistic differences in RNA editing regulation in SCZ between diverse ancestral populations, an aspect that needs to be investigated in the future.

We observed that differential editing sites were enriched in genes functionally important to the brain and mitochondria. Indeed, mitochondrial dysregulation has been implicated in SCZ by many studies[54,113–120]. Disruption in gene networks related to mitochondrial processes may lead to dendritic spine deficits and onset of SCZ symptoms[113]. In addition, abnormal SCZ brain connectivity may be associated with aberrant mitochondrial dynamics[116–120]. Finally, a reduced number of mitochondria was observed in certain layers of the neuronal somata and axospinous synapse terminals of SCZ brains[115]. Our findings further bolster the important implications of mitochondrial function in SCZ by demonstrating the alteration of expression or function of mitochondria-related genes by RNA editing.

We highlighted the function of two RNA editing sites located in the MFN1 gene, one of which is hypo-edited in SCZ. Using both mouse and human cells, we observed that the RNA editing recoding sites affect MFN1 function in mitochondrial fusion and cellular apoptosis. Previous work has shown that the ablation of MFN1 in mice leads to midgestational death, for which the embryonic fibroblasts display fragmented mitochondria due to significant reduction in mitochondrial fusion[104]. MFN1-mediated mitochondrial fragmentation has been suggested to lead to neurotoxicity in iPSCs[121]. In addition, low expression of MFN1 can lead to neuropathy, such as Charcot-Marie-Tooth disease, a neurodegenerative disease characterized by the demyelination of peripheral nerves[122]. To our knowledge, the literature on the potential impact of RNA editing on

mitochondrial function is very limited. For example, C-to-U editing by the APOBEC3G protein were predicted to suppress mitochondrial respiration relative to glycolysis in HuT78 cells[123]. However, the specific editing sites involved in such functions remain unknown. Thus, our study fills a significant gap by connecting RNA editing to important mitochondrial processes.

Mitochondria are highly dynamic and a balance between mitochondrial fusion and fission is important to brain development and function[124,125]. We observed a ~50% RNA editing level at the C-to-T recoding site and ~20% editing at the A-to-G site of MFN1 in normal brain samples (Fig. 2-4a), which appears to be conserved in mice (Supplementary Fig. 2-8a). We hypothesize that this intermediate level of RNA editing is important in maintaining the balance between mitochondrial fusion and fission. Higher or lower RNA editing may disrupt this balance and alter mitochondrial morphology and function. We observed that the edited versions of MFN1 generally led to reduced mitochondrial fusion, relative to the unedited protein, in both human and mouse cells, highlighting the functional relevance of the two RNA editing sites. However, it remains to be determined whether the reduction of RNA editing of the C-to-T site in SCZ patients from the nominal 50% causes mitochondrial fragmentation as reported in previous studies of SCZ brains[54]. Alternatively, it's possible that reduced MFN1 editing in SCZ is a way to compensate for mitochondrial fragmentation in the disease. In vivo experiments in mice or other animal models are needed to address this question.

Together, this work presents explicit evidence of the robust hypoediting trend in Europeans with SCZ, and supports the functional relevance of RNA editing in nuclear encoded mitochondrial genes. Although we cannot determine the contribution of RNA

editing to disease causality, we observed many functional pathways harboring various editing events that may contribute to disease risk (e.g., synaptic transmission and glutamate receptor activity). The large set of de novo editing sites resulting from our study will allow further elucidation of RNA editing differences observed in different ancestral groups. Taken together, our study provides an extensive characterization of RNA editing in SCZ and supplies valuable insight into the roles of dysregulated editing in mitochondrial function.

2.5 METHODS

2.5.1 RNA-seq data sets from PsychENCODE

RNA-seq data from the brain frontal cortex (FC) or dorsolateral prefrontal cortex (DLPFC) region were extracted from four cohorts in the PsychENCODE consortium: BrainGVEX, CMC, HBCC, and LIBD[94] (Supplementary Table 2-1). We followed strict quality control procedures to remove sample outliers in RIN, PMI, age, and other biological and technical variables[41]. Retained SCZ and control groups did not differ significantly in any biological or technical covariables (Supplementary Fig. 2-2, 2-4). The quality control procedure was conducted for European and African American samples separately in the HBCC and LIBD cohorts. All downstream analyses were conducted on each cohort and ethnicity separately in order to validate significant findings between cohorts and to limit noise due to potential cohort-specific batch effects. Final cohort sizes and number of case and control samples are provided in Supplementary Table 2-1.

2.5.2 Identification and annotation of RNA editing sites

RNA-seq was mapped using hisat2[126] to the human reference genome with default parameters, except allowing no mixed or discordantly mapped reads (--no-mixed, --no-discordant). Only uniquely mapped read pairs were used for downstream analysis. Unmapped reads can often be generated from regions with clusters of editing sites (i.e., "hyper-edited" regions) that fail to map due to copious mismatches. We applied a previously developed method to alleviate this issue[41,87]. Briefly, all adenosines in unmapped reads and the reference genome were converted to guanosines. This was followed by hisat2 alignment as described above, and then restoration of the original adenosines in the reads. Uniquely mapped reads from this step were combined with the originally mapped reads and used in the following RNA editing analysis.

RNA editing sites were identified using methods previously developed by our group[84,127,128]. Several filters were applied to remove loci resulting from spurious read mapping or sequencing errors[127]. Editing sites were supported by at least five samples, in which each was required to have at least two edited reads and five total reads. The loci occurring in at least 10% of samples within a cohort were labeled as "common" RNA editing sites, and used for downstream differential analysis (see below). Annotations of the genomic regions and host genes of RNA editing sites were obtained with the BiomaRt R package and Ensembl gene annotation[129,130]. The ANNOVAR[131] software was used to label the functional categories of RNA editing sites. Finally, editing sites were overlapped with ALU regions from RepeatMasker[132].

2.5.3 Differential RNA editing analysis

Differential RNA editing analysis was conducted for each cohort separately. To this end, we used a method previously developed in our lab that uncovers sites with either (1) significantly different average editing levels between SCZ and controls, or (2) differential editing prevalence between the two conditions[41]. As described below, this method adopts a strategy to allow a flexible read coverage requirement for each editing site, in order to adapt to the different total read coverage available to specific sites[41].

For each editing site e_i, we first identified the highest coverage possible (between a coverage of ≥20, 15, or 5 reads) at which there were a minimum of 5 samples per condition. After the highest possible read coverage *C* for e_i was chosen, we calculated separate average editing level per condition (A_{iSCZ} and $A_{iControl}$) utilizing samples with a minimum of *C* coverage. We then considered samples that fulfilled lower read coverage thresholds (≥15, 10, or 5), and included these samples in A_{iSCZ} or $A_{iControl}$ only if their inclusion did not alter the average editing level by > 0.03. A Wilcoxon rank-sum test was conducted to detect the difference in editing levels between SCZ and control groups. If an initial read coverage requirement *C* was not reached, then we tested all samples where e_i had ≥5 read coverage for at least 20% SCZ and 20% control samples. Differential editing sites were those with Wilcoxon rank-sum $P < 0.05$ and an effect size > 5%.

To identify differential editing sites that had significantly different editing prevalence between SCZ and controls, we used a Fisher's Exact test to compare the total numbers of SCZ and control samples with non-zero editing level vs. zero editing level, as previously described in Tran et. al.[41]. The adaptive procedure for minimum read coverage

requirements described above was also applied here. Editing sites with differential prevalence between the groups were those with $P < 0.05$ and effect size > 5%.

Differential RNA editing average (DREA) was calculated per sample using the mean editing level across all differential editing sites with read coverage ≥ 5 reads. A Wilcoxon rank-sum test was conducted to determine the significance of overall DREA trend between SCZ and control. Heatmaps of differential editing levels throughout the study were generated with the R package gplots[133].

2.5.4 Weighted gene co-expression network analysis (WGCNA)

We detected clusters of loci whose editing levels associated with SCZ diagnosis using the WGCNA package[95] in R for the BrainGVEX and CMC cohorts separately. We focused on common editing sites and those that showed large variations across samples. Specifically, 1) the site must be labeled as “common” in the cohort (defined above), 2) the site must have ≥ 5 reads in ≥ 90% of samples, and 3) the editing levels of the site must have a standard deviation > 0.1 across samples. WGCNA was conducted using automatic network construction and module detection, for which the soft threshold powers were set to be 2 or 3 for the BrainGVEX or CMC cohorts, respectively, to fit a scale-free topology[95].

The WGCNA network construction and consensus modules were chosen using the blockwiseConsensusModules function with default parameters, except for the soft threshold powers mentioned above. The eigengenes of each module were correlated with condition to identify the module most relevant to disease status. This correlation was conducted using a linear regression model, where biological (age and gender) and technical covariates (RIN, PMI, and total reads) were included. The largest module (the

turquoise module) significantly associated with disease status in both BrainGVEX and CMC (Supplementary Fig. 2-7).

2.5.5 Gene ontology (GO) enrichment of genes of interest

GO terms were obtained for a query gene set using the R package BiomaRt and Ensembl annotation[129,130]. For each query gene, we randomly picked a control gene with matched gene expression and gene length (±10% relative to that of the query gene). The controls of all query genes constitute one set of control genes. This process was repeated 10,000 times. Query genes without a matched control were excluded from the analysis. For each GO term, the number of its occurrences in the 10,000 sets of random controls was fit into a Gaussian distribution. The frequency of the term in the query gene set was then compared to this distribution to obtain an enrichment P value. Only terms that contain at least two genes in the query were considered. In rare cases, a particular GO term does not occur in the control sets. Its p value was set to $1e^{-100}$ for visualization purposes.

2.5.6 Cell type proportion from bulk RNA-seq

FPKM for bulk RNA-seq was calculated based on read counts per gene obtained using the HTSeq software[134] and total mapped reads from hisat2[135]. Cell type proportion for five main cell types in the human brain (neurons, astrocytes, oligodendrocytes, microglia, and endothelial cells) were estimated using the “impute cell fractions” method in the CIBERSORTx software[92]. CIBERSORTx was provided with bulk gene expression and a cell type signature matrix of genes, which was derived from single-cell RNA-seq of human prefrontal cortex (PFC) in previous studies[96,136]. After obtaining normalized cell

type proportion values for each sample and cell type, we conducted Wilcoxon rank sum tests between conditions for each cell type to determine differences in proportion. Cell types with FDR adjusted $p < 0.05$ were marked with red asterisks (Supplementary Fig. 2-1i and Supplementary Fig. 2-5b).

2.5.7 Cell type-specific differential RNA editing average (DREA)

Cell type specific DREA was calculated by taking the average editing level of all differential editing sites located in the signature genes of a cell type. Differential editing sites with at least 5 reads were included. A Wilcoxon rank sum test was conducted between conditions to ascertain cell type DREA differences (red asterisk denoting FDR adjusted $p < 0.05$, Supplementary Fig. 2-1j, Supplementary Fig. 2-5c).

2.5.8 Single-nucleus RNA-seq data and RNA editing quantification

We obtained previously published single-nucleus transcriptomes of six major brain cell types – neurons (both excitatory and inhibitory), astrocytes, oligodendrocytes, microglia, endothelial, and oligodendrocyte progenitor cells – from the PFC of 24 control subjects[91]. Data of the same cell type were combined for RNA editing detection and transcripts per million (TPM) calculation, which were conducted similarly as described above.

For each cell type in each sample, we estimated the fraction of genes expressed with TPM ≥ 5 among all differentially edited genes in the BrainGVEX cohort (Fig. 2-1g). In addition, we estimated the fraction of differentially edited sites in the BrainGVEX cohort that have detectable editing in each cell type. That is, for each cell type in each sample,

we obtained the ratio between the number of detected RNA editing sites and the total differential editing sites in BrainGVEX (Fig. 2-1h).

2.5.9 MPRA of 3' UTR differential editing sites

A total of 770 differential editing sites located in 3' UTRs from the BrainGVEX, CMC, HBCC, or LIBD data were included in the MPRA experiment in HEK293 cells. Specifically, we synthesized 200nt-long oligos (Twist Biosciences) containing cloning adaptors and 158nt-long test sequences with the editing site at the center of the test region. If the editing site is close to either boundary of the 3' UTR, the flanking regions of the editing site were adjusted such that the 158nt-long test sequence resided within the 3' UTR. The edited (G) and unedited (A) versions of the editing sites were both synthesized. The test sequences were cloned into the 3' UTR of the *eGFP* gene in the master plasmids. The plasmid library was then electroporated into HEK293 cells followed by RNA extraction 24 h post cell transfection. Finally, the test sequences were amplified from both plasmid library and mRNA to generate DNA-seq and RNA-seq libraries, similarly as in previous MPRA experiments[137,138]. Sequencing data of the plasmid DNA and mRNA were compared to identify sites associated with significant expression differences between the two alleles (A and G) using MPRAnalyze[139]. FDR ≤ 0.1 and $|\ln(\text{Fold Change})| \geq 0.1$ were required to call significance.

2.5.10 Correlation between 3'UTR editing and gene expression

For each differential editing site in the 3' UTR, we correlated the editing levels and the expression levels of the host gene (log FPKM) across all samples, using a linear

regression model that included sample PMI, RIN, age of death, and sex as covariates. Only sites with ≥ 5 reads in ≥ 10 samples were deemed to be testable. This analysis was conducted for each of the four cohorts separately. However, FDR correction was applied to p values obtained from all cohorts combined. The t-statistic, the number of standard deviations by which the correlation deviated from zero, was used to determine association directionality for visualization (Fig. 2-3c).

2.5.11 Gene regulatory network analysis

We used GeneMANIA[98] to obtain known genetic interactions for 195 genes whose 3' UTR differential editing sites were significantly associated with their respective gene expression. GeneMANIA identified 20 additional genes strongly connected with the 195 genes. The gene regulatory network encompassing 215 genes was visualized using Cytoscape v3.8.2[99]. GO enrichment was conducted for the 195 genes to identify functional categories. Genes that fell into similar categories were grouped together in the network visualization. FDR corrected p values (q values) for the GO categories were labeled on the network. For those categories with multiple GO terms, the most significant q value was shown (Fig. 2-3e).

2.5.12 MFN1 editing in zebrafish

RNA-seq extracted from zebrafish (*Danio rerio*) brain tissue was obtained from Wong et. al. using NCBI's Gene Expression Omnibus database (GSE61108)[103]. The neurotranscriptome profiles for a total of four strains and 160 samples were analyzed. Same-sex individuals were pooled into a biological replicate. A total of 16 RNA-seq

datasets (2 biological replicates from each sex for each strain) were mapped to genome assembly GRCz11 using the hisat2[126] software as described previously. Editing ratios for orthologous MFN1 sites (mfn1b in zebrafish) were manually calculated based on the integrative genomics viewer (IGV)[140].

2.5.13 Cell culture

Mfn1 and *Mfn2* double knockout mouse embryonic fibroblasts (MEF) were kindly gifted by Dr. David Chan (California Institute of Technology). HEK293 cells were kindly gifted by Dr. Jing Huang (UCLA). Mouse neuro2a (N2a) and human embryonic kidney (HEK293T) cells were obtained from ATCC. Cells were maintained with Dulbecco's modified eagle medium (Gibco, 10569010) supplemented with 10% fetal bovine serum (Gibco, 10082147) and 1x antibiotic-antimycotics (Gibco, 15240096) at 37 °C and 5 % CO_2. For the neuronal differentiation of N2a cells, the cells were seeded at 40 - 50% of density and grown for 24h in complete medium. Following this, the cells were washed with 1x Dulbecco's phosphate-buffered saline (Gibco, 14190144) and then replaced in serum-free medium.

2.5.14 MFN1 editing constructs

Human and mouse *MFN1* constructs were kindly gifted by Dr. Orian Shirihai (UCLA). MFN1 editing mutations were generated by the Q5 site directed mutagenesis kit (NEB, E0554S), then cloned into pqCXIP (Clontech, 631516) for stable Mfn1 expression in MEF cells.

2.5.15 RNA isolation, RT-PCR amplification and analysis of RNA editing

Cultured cells and brain tissues were homogenized in TRIzol (Thermo Fisher Scientific, 15596018). The mixture was incubated on ice for 15 min. Chloroform was added to the mixture and incubated at room temperature for 10 min. The mixture was centrifuged at 12,000x g for 15 min, and the top layer was carefully extracted. An equal volume of 100% ethanol was added to the top chloroform layer and mixed thoroughly. Total RNA of N2a cells was also extracted using the TRIzol reagent. RNA was further purified using the Direct-zol RNA MiniPrep Plus kit (Zymo Research, R2072) according to the manufacturer's protocol. Reverse transcription was performed on 1 µg total RNA for 1h at 42 °C using random hexamer primer and SuperScript IV (Thermo Fisher Scientific, 18090050). The cDNA product was detected by PCR using the *Mfn1* gene-specific primer set. Amplification was performed for 30 cycles, consisting of 30s at 95 °C, 30s at 55 °C, and 1 min at 72 °C. The products from RT-PCR were resolved on 1% agarose gels. The appropriate PCR product was excised and the DNA was extracted, purified and analyzed by Sanger sequencing. C-to-T editing levels were calculated as relative peak heights (that is, ratio between the T peak height and the combined height of C and T peaks: height T / (height C + height T)).

2.5.16 TOPO cloning and clonal sequencing

PCR products were run on 1% agarose gel and visualized under ultraviolet light. The band with the expected size was isolated by Zymoclean Gel DNA Recovery kit (Zymo Research, D4002) according to the manufacturer's protocol. The PCR product was inserted into kanamycin-resistant pCR 2.1-TOPO vector (Thermo Fisher Scientific,

450641). Ligated clones were transformed into One Shot TOP10 Chemically Competent *E. coli* (Thermo Fisher Scientific, C404003). Transformed cells were streaked on LB-agar plates containing kanamycin and X-Gal as selection markers and incubated overnight at 37 °C. Ten white colonies were randomly selected and each colony was inoculated overnight in LB containing kanamycin. Plasmid was extracted using Plasmid DNA Miniprep Kits (Thermo Fisher Scientific, K210011). Miniprep samples were subject to Sanger sequencing. The number of clones presenting a thiamine (T, representing C>U editing) or guanine (G, representing A>I editing) peak at the editing sites in the *Mfn1* gene was counted to determine the editing ratio.

2.5.17 Production of lentivirus and cell transduction for protein knockdown

Constructs containing pLKO1 non-target control shRNA (SHC016), Apobec2-targeting shRNA (TRCN0000112015) or Apobec3-targeting shRNA (TRCN0000197906) were used. We produced lentiviruses via co-transfection of pCMV-d8.91, pVSV-G and pLKO1 into HEK293T cells using Lipofectamine 3000 (Thermo Fisher Scientific, L3000015). Transduction was carried out according to the standard protocol of the ENCODE consortium. Briefly, viruses were collected from conditioned media after 48h co-transfection. Lentivirus-containing medium was mixed with the same volume of DMEM containing polybrene (8 μg/mL), which was used to infect N2a cells. After 24h, cells were incubated with 3 μg/mL puromycin for 24h. Knockdown efficiency was evaluated by real-time quantitative PCR (Bio-Rad).

2.5.18 Analysis of mitochondrial morphology

Mfn1 and *Mfn2* double knockout MEF cells were seeded and incubated for 24h in 384-well plates (E&K Scientific, EK-30091) and then stained with 200 nM MitoTracker Green FM (Invitrogen, M7514) for 30 min at 37 °C. Mitochondria were visualized with the Zeiss LSM 780 confocal microscope and ZEN software (Zeiss).

2.5.19 Propidium iodide staining

Cells were stained using 1 µg/mL Propidium iodide (Invitrogen, V13242) and 1 µg/mL Hoechst 33342 (Invitrogen, H3570) for 10 min. Images were obtained using confocal microscopy and analyzed by the ImageJ software (http://imagej.nih.gov/ij/).

2.5.20 Mitochondrial fractionation

Mitochondrial and cytosol fractions were isolated using the mitochondria isolation kit for cultured cells (Thermo Fisher Scientific, 89874). Briefly, 80% confluent cells in 10cm culture dish were resuspended in 800µl of reagent A and incubated for 2 min on ice. Then 10µl reagent B was mixed by vortexing for 5 min. Samples were mixed with 800µl reagent C and centrifuged at 700x g for 10 min at 4 °C. Supernatant was transferred to a new tube and centrifuged at 12,000x g for 15 min at 4 °C. The supernatant was then collected as the cytosol fraction. The mitochondrial fraction in the pellet was washed once with 500µl DPBS and centrifuged at 12,000x g for 5 min at 4 °C. The mitochondrial fraction was dissolved in 200ul DPBS with sonication at 25% amplitude for 10 s (twice) with an Ultrasonic Processor 120 W, 20KHz (Thermo Fisher Scientific).

2.5.21 Western Blot

Cells were lysed and separated via the Novex NuPAGE system (Invitrogen, NP0008, NP0001 and LC3675) and the ExpressPlus PAGE Gel (GenScript). PVDF membranes (Millipore, IPVH304F0) were used for transfer, and then probed with the following primary antibodies in 3% BSA / 0.1% TBST: β-Tubulin (Santa Cruz Biotechnology, sc-23949), ATP5a (Abcam, ab14748), cytochrome C (Santa Cruz Biotechnology, sc-13156), MFN1 (Abcam, ab57602), MFN2 (Abcam, ab56889), β-actin (Santa Cruz Biotechnology, sc-47778). The antibodies were detected using horseradish peroxidase-conjugated antibodies (Invitrogen, 31430), the ECL Prime Western Blotting System (GE Healthcare, RPN2232) and the Syngene Pxi Imager.

2.5.22 MFN1 genome editing

We introduced *MFN1* editing sites to the genome of HEK293T cells via Prime editing[106]. The spacer and extension sequences were designed according to the guidelines provided previously[106]. The designed oligos were cloned into the pU6-pegRNA-GG-acceptor (Addgene, 132777) to generate both pegRNA and nick gRNA expressing constructs (oligo sequences were listed in Supplementary Table 2-4). Plasmids expressing pegRNA (250ng), nick gRNA (83ng) and prime editor (750ng), namely pCMV-PE2 (Addgene, 132775), were co-transfected into HEK293T (7,500 cells/well in 48-well plates) with Lipofectamine™ 3000 Transfection Reagent (Thermo Fisher Scientific, Cat# L3000015) according to the manufacturer's protocol. After 72 hours, genomic DNA was extracted, amplified by PCR and sequenced via Sanger sequencing to confirm genome editing events. To optimize the editing efficiency, different

combinations of pegRNAs and nick gRNAs (PE2, PE3, and PE3b) were tested. The prime binding site (PBS) and reverse transcription (RT) template were also optimized to 9nt for PBS and 19nt for RT template. The optimal condition (PE3b, 9nt PBS, and 19nt RT template) was used in scaled-up experiments to generate single-cell clones with MFN1 genome editing by serial dilution.

2.5.23 Mitochondria hyperfusion test

HEK293T cells were seeded on the 4-chambered coverglass (Nunc, Lab-TekII). 10 μM of Cycloheximide (CHX) was added to 70% confluent cells for 30 min and then stained with MitoTracker™ Green FM (Invitrogen, M7514) for 30 min at 37 °C. Mitochondria images were obtained using the Leica DMI-4000 confocal microscope and Leica application suite software. The mitochondrial hyperfusion score was calculated as described previously[107].

2.5.24 Statistics for MFN1-related experimental data

GraphPad Prism 9 was used for statistical analysis and graphical display of the data. All graphs represent mean ± standard error of the mean (SEM). Statistical tests and significance values were indicated in the figure legends.

2.6 ACKNOWLEDGEMENTS

We thank Drs. David Chan (California Institute of Technology) and Orian Shirihai (UCLA) for sharing the *Mfn1/2* knockdout MEF and wildtype cells and the *Mfn1* overexpression constructs. We thank Dr. Jing Huang (UCLA) for sharing HEK293 cells. We appreciate the helpful discussions with Drs. Carla Koehler and Alexander Van der

Bliek. We thank N. Prunet and the MCDB/BSCRC Microscopy Core for training and microscope facilities. We thank members of the Xiao laboratory for helpful discussions and comments on this work. This work was supported in part by grants from the National Institutes of Health (R01MH123177, R01AG075206 to X.X., and R01AG037514 and RF1AG049157 to D.W.W.). M.C. and T.W.C. were supported by the NIH T32LM012424. T.F. was supported by the UCLA Hyde Fellowship and Dissertation Year Fellowship. K.A. was supported by the University of California-Historically Black Colleges and Universities (HBCUs) Fellowship.

2.7 AUTHOR CONTRIBUTIONS

M.C., J.H.B. and X.X. designed the study with inputs from all other authors. T.F., J.H.B., H.I.J., and S.P. conducted the molecular, cellular and biochemical experiments. M.C., T.F., T.W.C. and K.A. conducted the bioinformatics analyses. X.X., D.W.W. and J.H.B. provided supervisory inputs. All authors contributed to the writing of the paper. All authors approved the final manuscript.

2.8 COMPETING INTERESTS

The authors declare no competing interests.

2.9 FIGURES

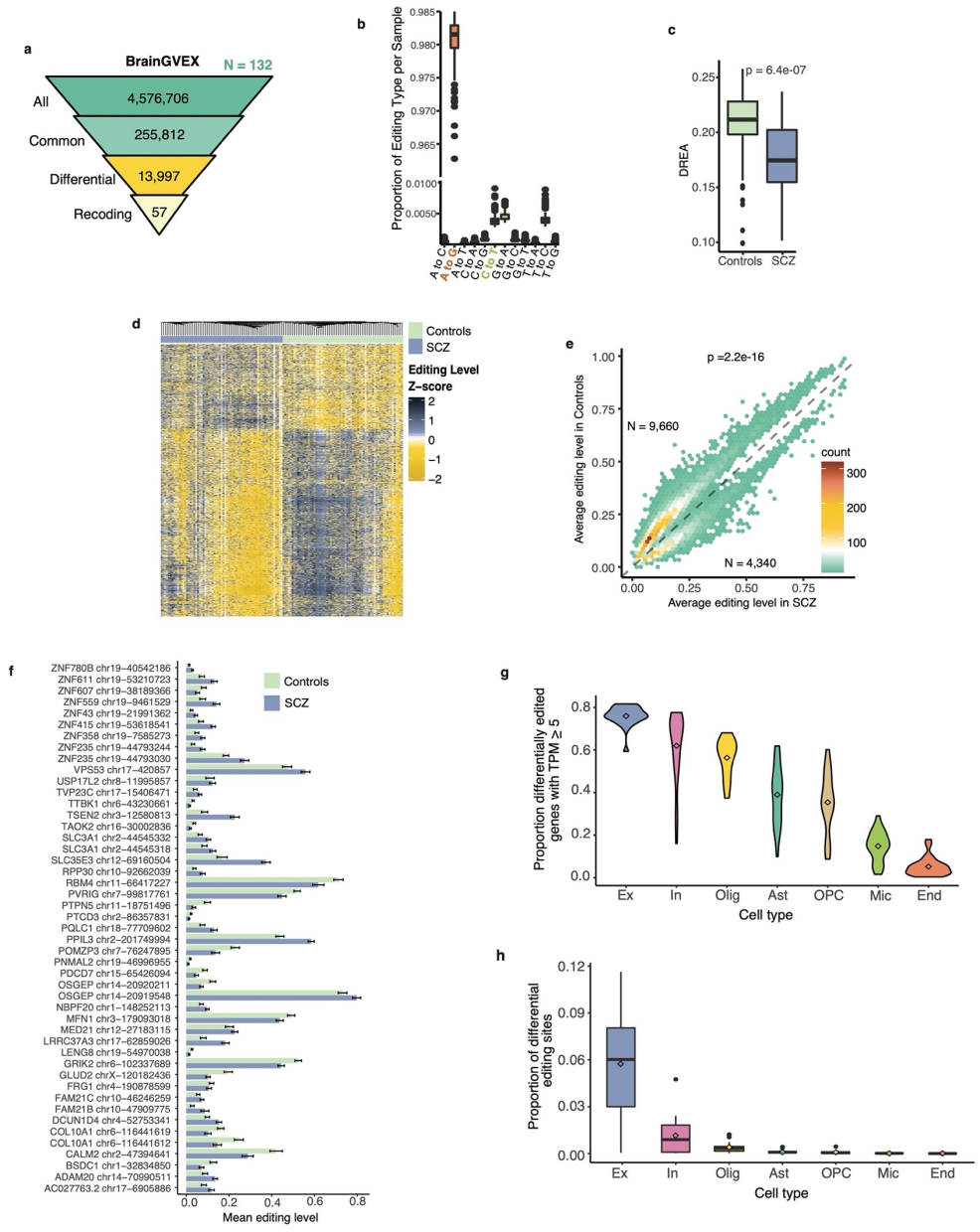


Figure 2-1. Overview of RNA editing analysis in the BrainGVEX cohort.

a, Summary of editing sites detected in each step of the RNA editing analysis. Top to bottom: All detected RNA-DNA differences (RDDs), common sites with non-zero editing in ≥10% of samples, sites differentially edited between SCZ and control (see Methods), and nonsynonymous protein recoding sites among the differential sites. N=132: total number of samples included in the analysis. **b,** Proportion of each type of RDD among common sites detected per

sample. **c,** Differential RNA editing averages (DREA) per sample separated by condition for all differential sites covered by least 5 total reads. P value was calculated via Wilcoxon rank sum test. **d,** Hierarchical clustering of differential editing sites (rows) and samples (columns). Z-scores were calculated for each site across all samples. **e,** Average editing levels of differential editing sites in SCZ and controls. Numbers (N) of editing sites that were up- or down-regulated in SCZ are shown, which were compared via Chi-squared test (p value shown at the top). **f,** Editing levels in controls or SCZ of each differential protein recoding site. Error bars correspond to the standard error of mean (SEM). **g,** Proportion of genes with ≥ 5 TPM in each cell type among all differentially edited genes in the single-nucleus RNA-seq data. Ex: excitatory neurons, In: inhibitory neurons, Olig: oligodendrocytes, Ast: astrocytes, OPC: oligodendrocyte progenitor cells, Mic: microglia, End: endothelial cells. **h,** Proportion of observed editing sites in each cell type among all differentially edited sites. Cell type abbreviations are the same as in **g**.

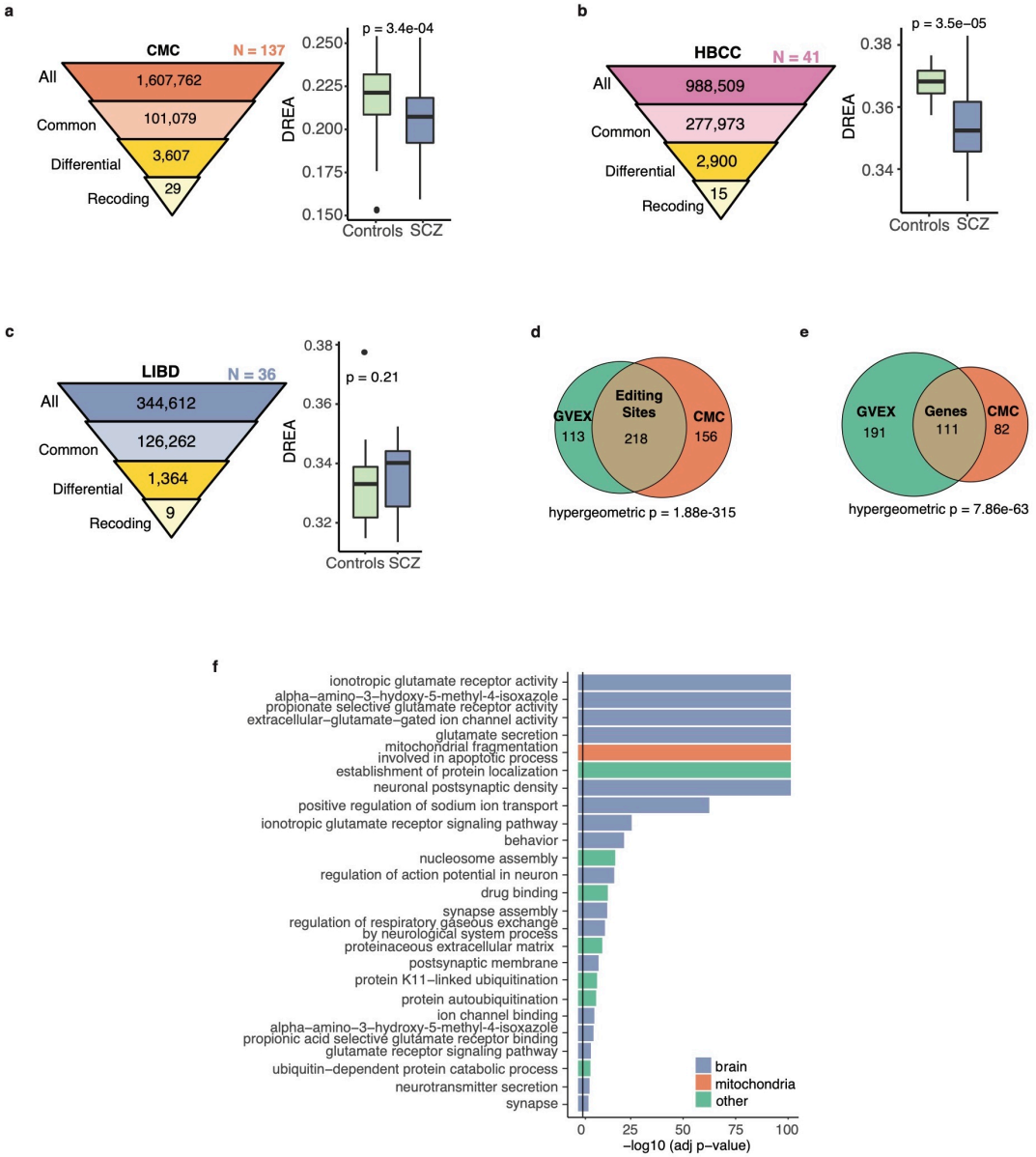


Figure 2-2. Comparison of RNA editing across cohorts.

a-c. Summary of editing sites detected in each step of the RNA editing analysis for the CMC, HBCC and LIBD cohort, respectively, similar as Fig. 1a and 1c. **d,** Overlap between RNA editing sites in the turquoise modules resulting from WGCNA of the BrainGVEX and CMC cohorts. Only sites testable for both cohorts are displayed. P value was determined via the hypergeometric test. **e,** Similar to **d** but for genes harboring editing sites in the turquoise modules. **f,** GO enrichment analysis of genes shared by the BrainGVEX and CMC analysis in **e**.

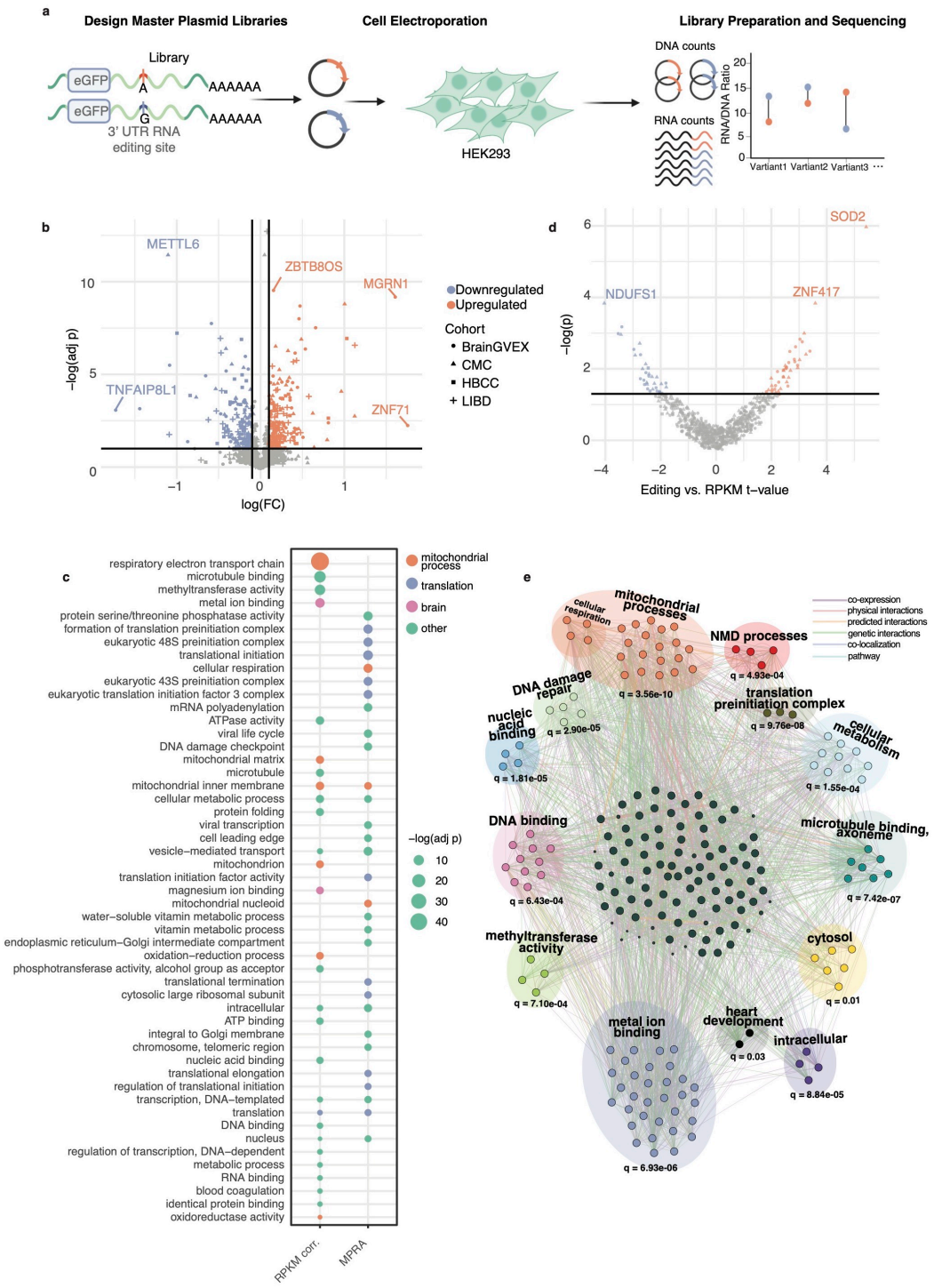


Figure 2-3. 3' UTR editing associated with gene expression.

a, Schema of the MPRA experiment to identify functional 3' UTR editing sites that alter gene expression. **b,** Expression fold change (FC) and adjusted p value (see Methods) of 3' UTR editing sites included in the MPRA. Purple and orange dots correspond to editing sites that significantly alter gene expression (FDR ≤ 0.1 and |ln(Fold Change)| ≥ 0.1). **c,** GO

enrichment analysis of genes containing 3' UTR differential editing sites associated with gene expression through experimental (MPRA) or bioinformatic (RPKM correlation) analysis. **d,** Significance (p value) and directionality (t-statistic, jittered for visualization) of the correlation between editing levels of 3' UTR differential editing sites and their respective gene expression levels. Orange and purple colors denote $p < 0.05$. **e,** Gene network constructed by GeneMANIA[98] for the union of significant genes in **b** and **d**. Genes are grouped into biological categories based on GO enrichment analysis. The most significant q-value for the overarching GO category is displayed.

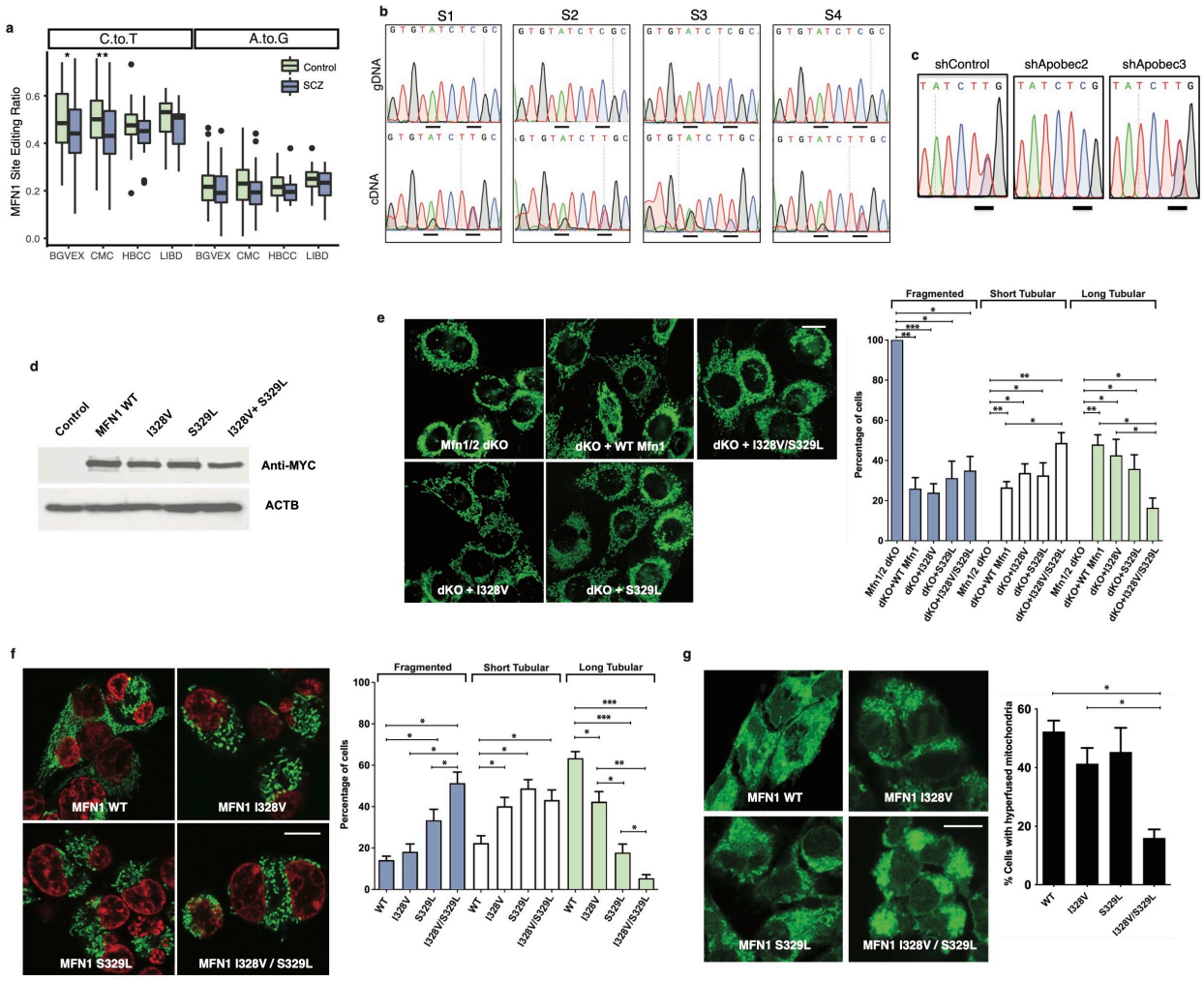


Figure 2-4. Functional characterization of MFN1 editing in mouse and human.

a, MFN1 C-to-T and A-to-G editing levels per cohort for SCZ and Control samples. *p < 0.05, **p < 0.001, Wilcoxon rank sum test. **b,** Experimental validation of the C-to-T and A-to-G MFN1 RNA editing sites in four human brain samples (S1-S4). Sanger sequencing traces of genomic DNA (top) and cDNA (bottom) are shown, with the A-to-G (left) and C-to-T (right) editing sites underlined. **c,** Experimental testing of the C-to-T editing site (underlined) in mouse N2a cells with control shRNA (shControl) or shRNA targeting Apobec2 or Apobec3, respectively. Sanger sequencing traces of cDNAs are shown**. d,** Stable expression of wildtype (WT) Mfn1 and editing mutants (Mfn1 I328V, S329L, I328V/S329L) (with Myc tags) in Mfn1/2 dKO MEF cells measured by Western blot. Control: empty vector expression. **e,** Mitochondrial morphology in Mfn1/2 dKO MEF cells with stable expression of WT and mutant Mfn1/2 as shown in **d.** Mitochondria was stained by 100nM MitoTracker (green) for 30 min. Scale bar, 10 μm. Bar plots show quantification of mitochondrial morphology of 120 cells in 3 biological replicates. Fragmented mitochondria (blue), short tubular mitochondria (white), long tubular mitochondria (green) were quantified separately (*p < 0.05, **p < 0.001, ***p<0.0001, unpaired two-tailed

t-test). Error bars show ±SEM. **f,** Mitochondrial morphology in HEK293T cells with WT MFN1 or MFN1 editing mutants (MFN1 I328V, S329L, I328V/S329L). Green: MitoTracker staining for mitochondria, red: Hoechst staining for nucleus. Scale bar, 10 μm. Bar plots show quantification of mitochondrial morphology of 100 cells in 3 biological replicates, similarly as in **e**. (*p < 0.05, **p < 0.001, ***p<0.0001, unpaired two-tailed t-test). Error bars show ±SEM. **g,** Mitochondrial morphology of WT and mutant HEK293T cells treated with cycloheximide for 30 min. Mitochondria were stained by MitoTracker (green). Scale bar, 10 μm. Bar plots indicate cell counts with hyperfused mitochondria (*p < 0.05, unpaired two-tailed t-test). Error bars show ±SEM.

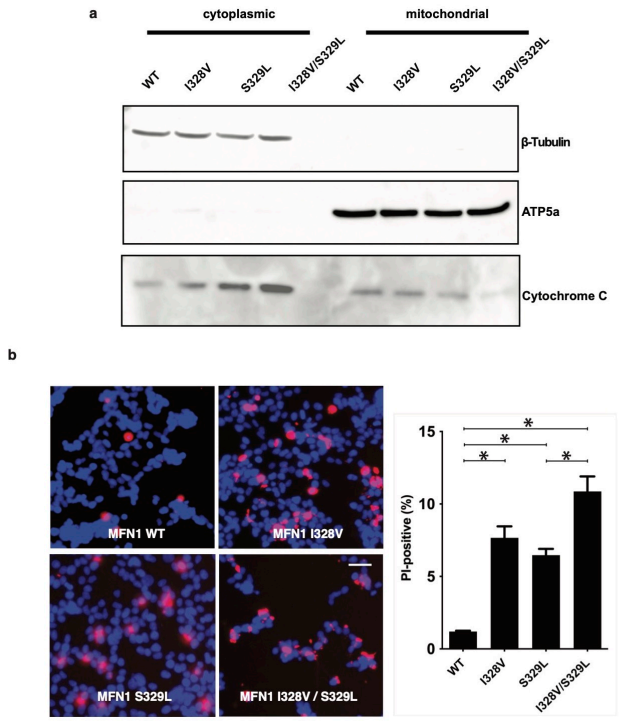
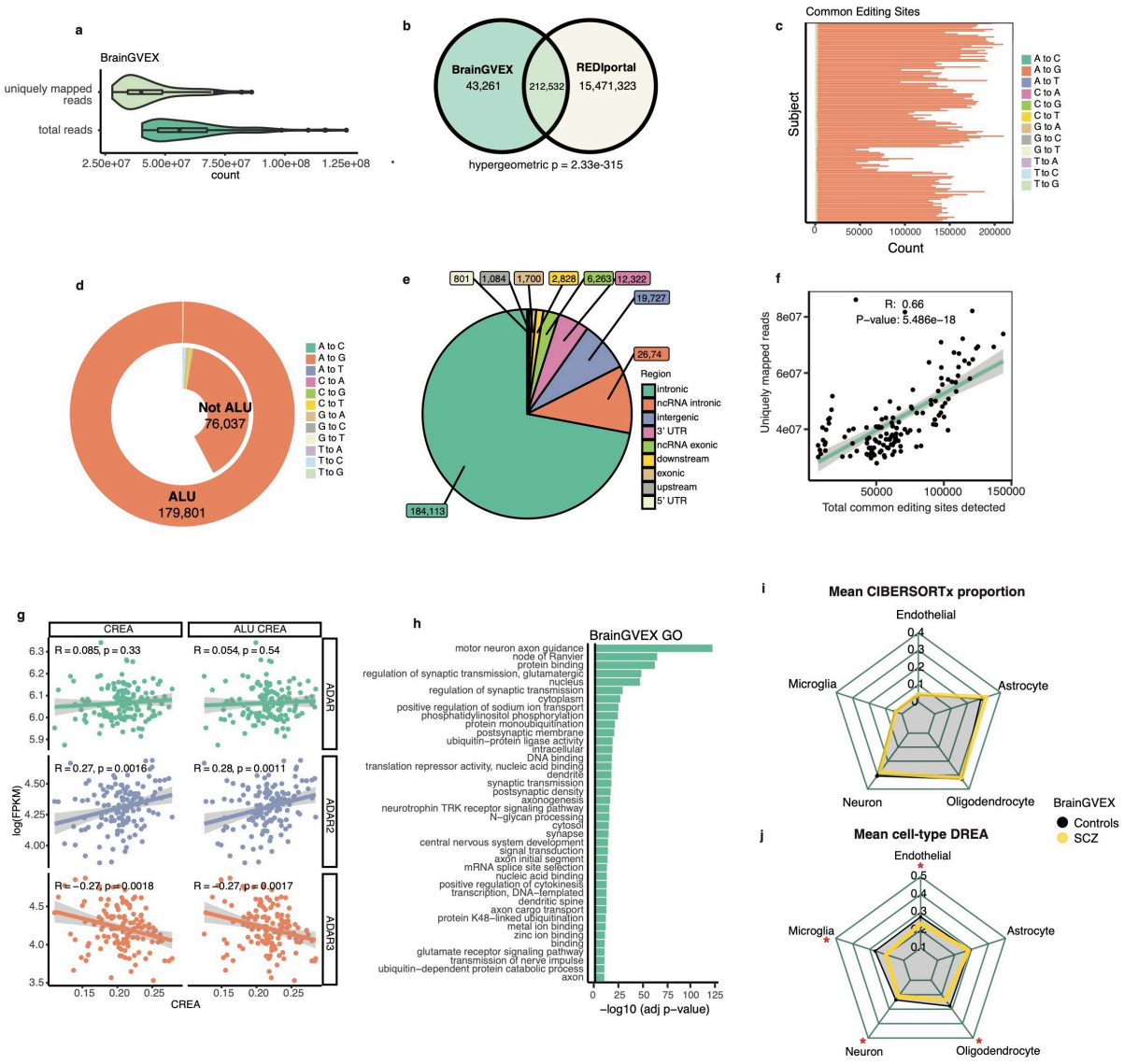


Figure 2-5. The effect of MFN1 editing on cytochrome C release and apoptosis.

a, Western blot of Cytochrome C, β-Tubulin (cytoplasmic marker) and ATP5a (mitochondrial marker) in mitochondrial and cytoplasmic fractions of HEK293T cells (WT and MFN1 mutants). **b,** PI staining of HEK293T cells (WT and MFN1 mutants) to measure cytotoxicity. Scale bar, 10 μm. Bar plots show % of dead cells (*$p < 0.05$, unpaired two-tailed t-test). Error bars correspond to ±SEM.

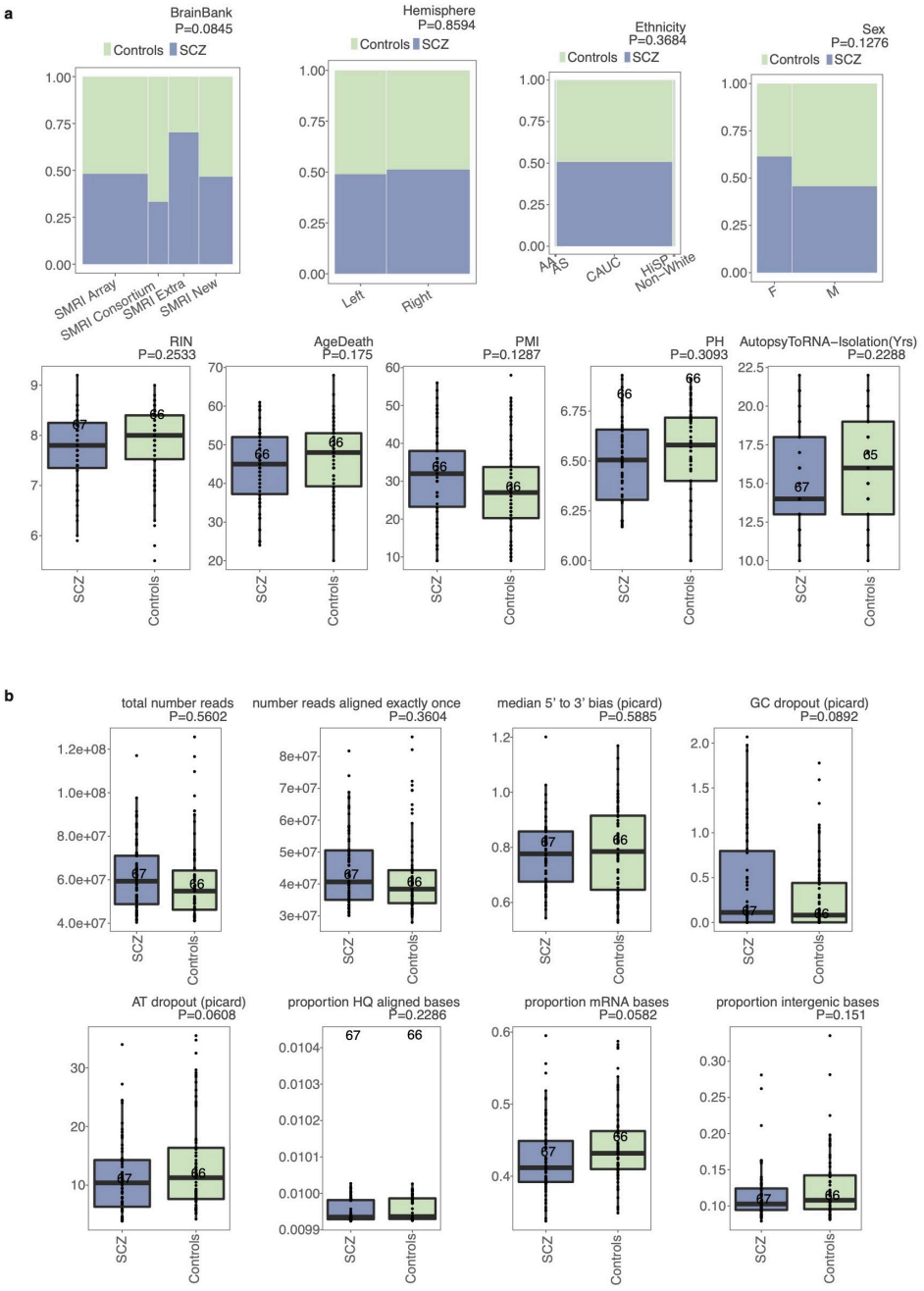
2.10 SUPPLEMENTARY FIGURES



Supplementary Figure 2-1. Additional characterization of BrainGVEX editing sites.

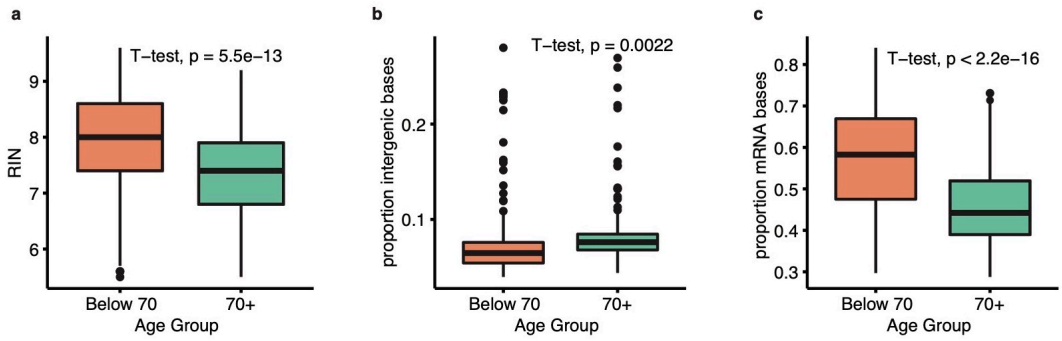
a, The number of total reads or uniquely aligned reads for the BrainGVEX RNA-seq samples. **b,** Overlap between RNA editing sites detected in the BrainGVEX cohort and those cataloged by the REDIportal database. P value was calculated via the hypergeometric test. **c,** Total count of common sites detected per sample grouped by RDD type. **d,** Donut plot of common sites in ALU regions (outside) and non-ALU regions (inside) grouped by RDD type. **e,** Distribution of common sites in different types of genomic region. **f,** Pearson correlation between the number of uniquely mapped

reads and the total number of common sites detected per sample. **g,** Pearson correlation between the expression levels of ADARs and common RNA editing average (CREA) for all common sites (left) or common sites in ALU regions (right). **h,** GO enrichment results for all differentially edited genes in the BrainGVEX cohort. **i,** Average cell type proportions in control and SCZ samples determined via CIBERSORTx. No significant difference was observed (FDR > 0.05, Wilcoxon rank sum test). **j,** Average DREA for control and SCZ samples calculated for differential sites in the signature genes of each cell type established in Yu et. al[93]. Red asterisks: FDR < 0.05 comparing SCZ and controls by Wilcoxon rank sum test.



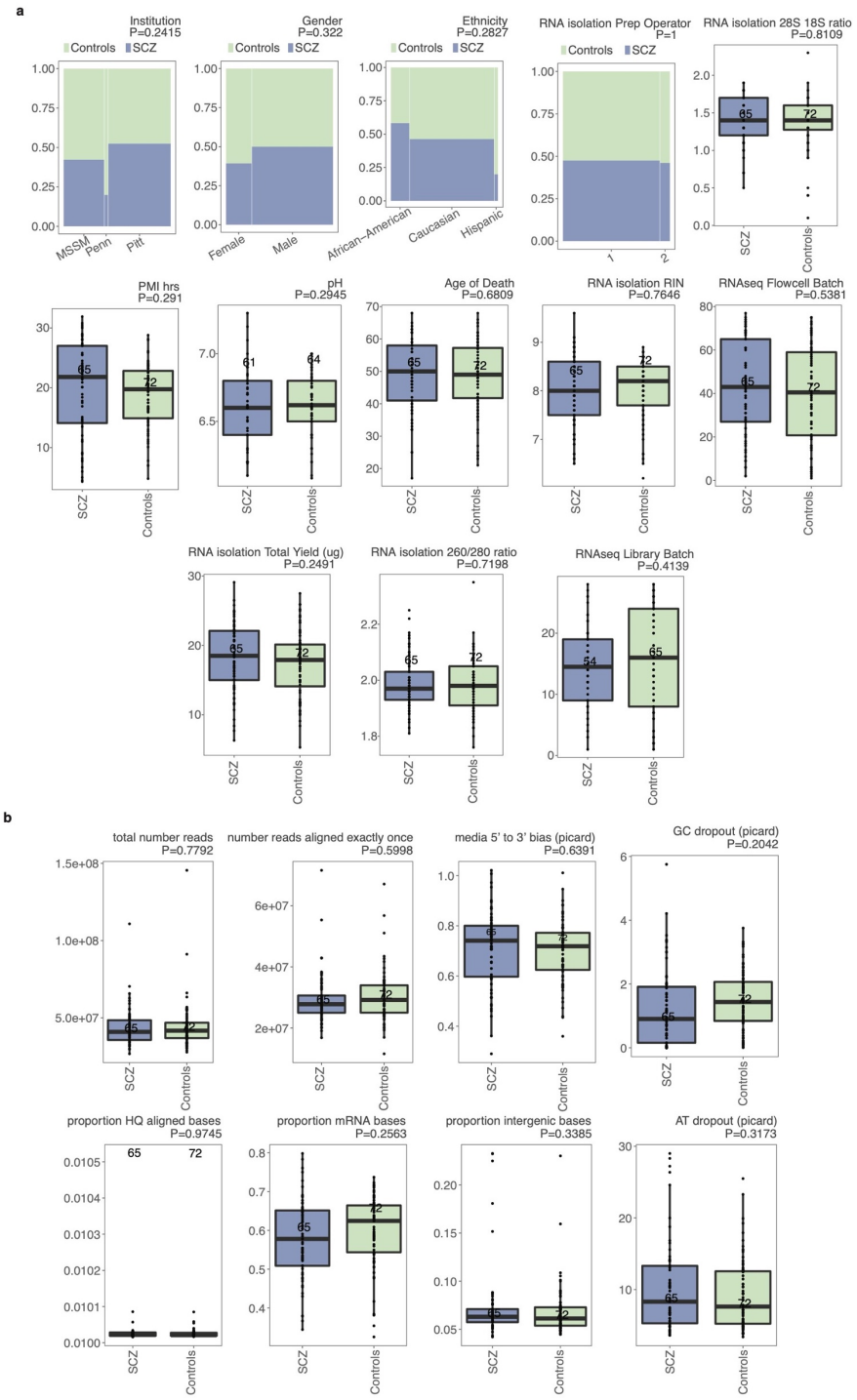
Supplementary Figure 2-2. Comparison of potential confounding variables between SCZ and controls of the BrainGVEX cohort.

P values were calculated via Pearson correlation and Fisher's exact test for numeric and categorical covariates, respectively. **a,** Biological and technical covariates for SCZ and controls after completing the sample QC procedure (Methods). **b,** RNA-seq mapping metrics for SCZ and controls after completing the sample QC procedure (Methods).



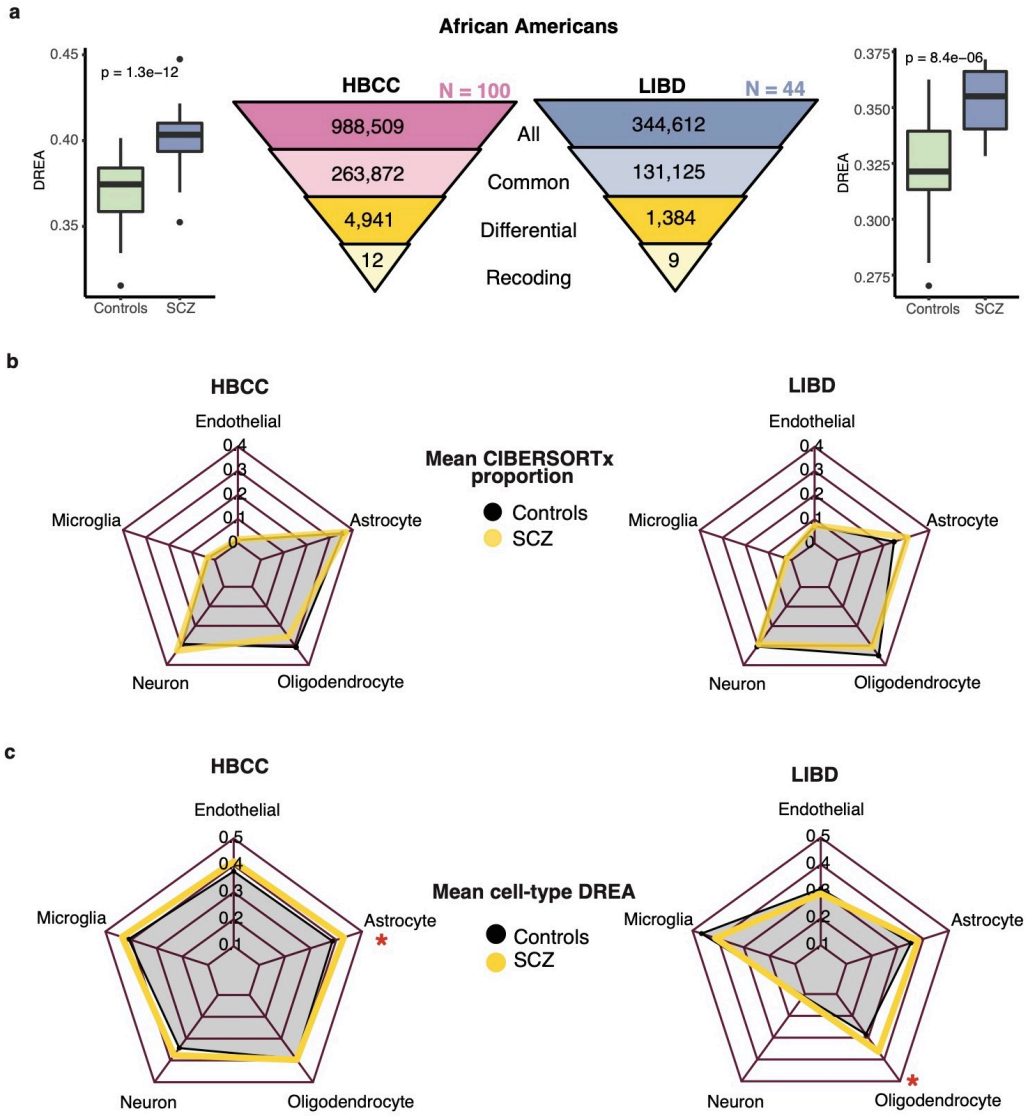
Supplementary Figure 2-3. Meta data for CMC individuals below and above 70 years of age.

Distribution of RIN **(a)**, proportion of bases mapped to intergenic regions **(b)** and proportion mapped mRNA bases **(c)** in the two age groups.



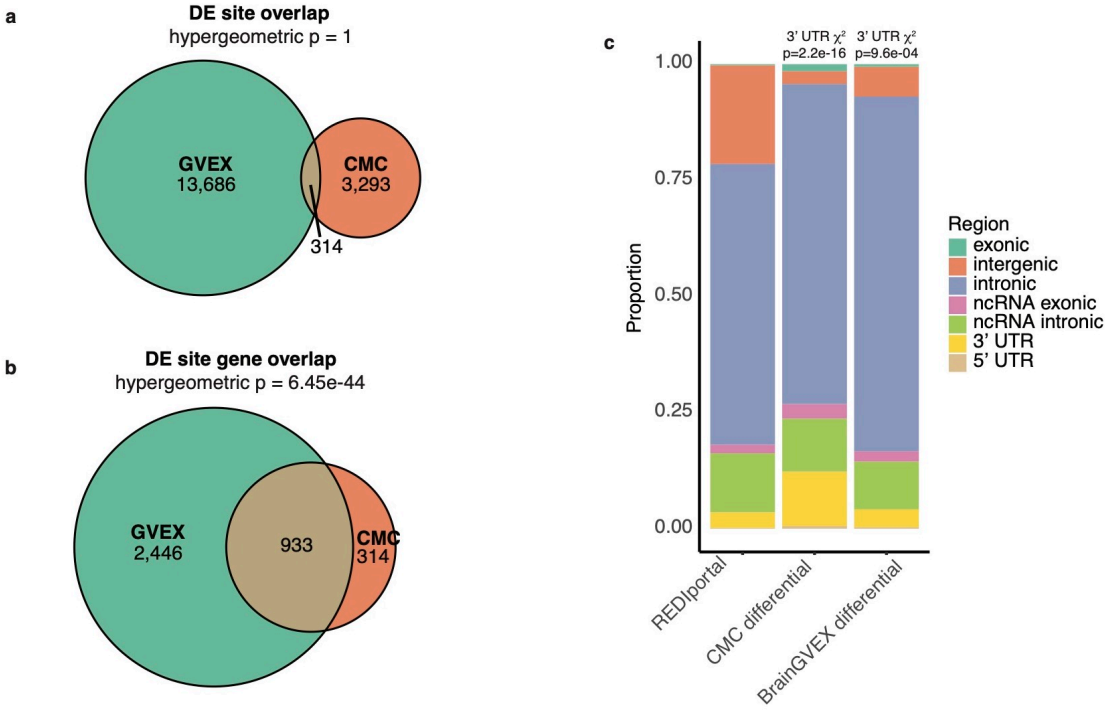
Supplementary Figure 2-4. Comparison of potential confounding variables between SCZ and controls of the CMC cohort.

Similar to Fig.S2-2, for CMC samples (<70 years of age) after completing the QC procedure for samples below 70 years of age.



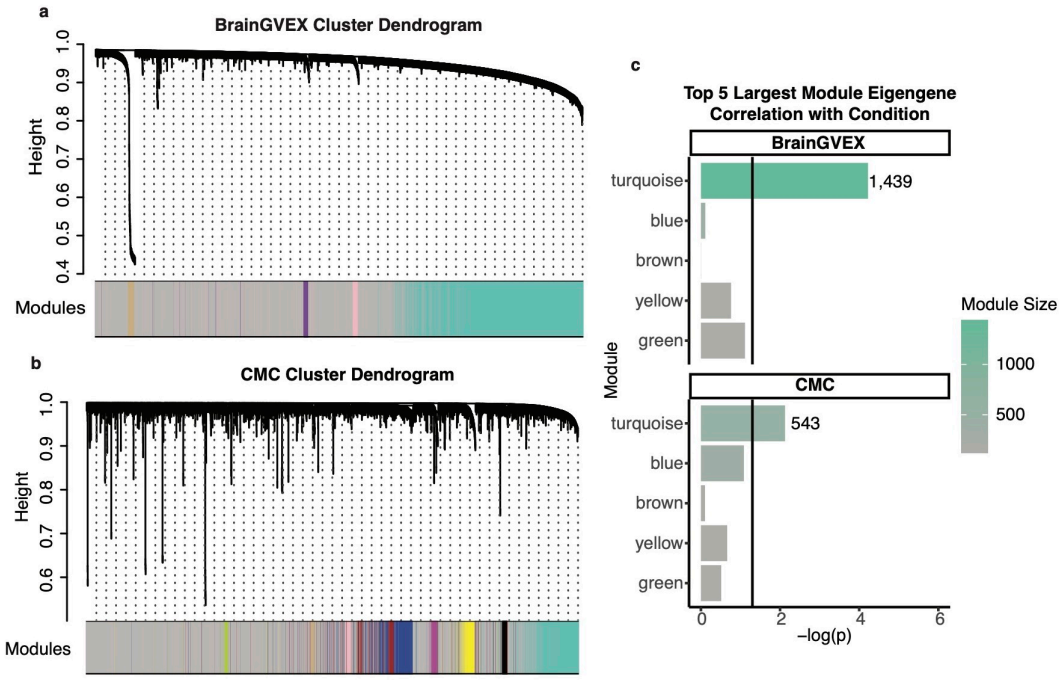
Supplementary Figure 2-5. RNA editing in African American samples of the HBCC and LIBD cohorts.

a, Overview of all, common, differential, and recoding sites observed in each cohort. Boxplots show sample DREAs separated by condition. P values were calculated via Wilcoxon rank sum test. **b,** Average cell type proportions in SCZ and controls determined by CIBERSORTx. No significant difference was observed (FDR > 0.05, Wilcoxon rank sum test). **c,** Average DREA of SCZ and controls calculated for differential sites in the signature genes of each cell type established in Yu et. al[93]. No differential editing sites were detected in neuronal genes in the LIBD cohort. Red asterisks: FDR < 0.05 between SCZ and controls, Wilcoxon rank sum test.



Supplementary Figure 2-6. Overlap of BrainGVEX and CMC and 3' UTR enrichment.

a, Overlap between differential editing sites in the BrainGVEX and CMC cohorts. P value was calculated via hypergeometric test. **b,** Similar to **a**, for differentially edited genes. **c,** Distribution of REDIportal sites, differential sites in CMC or BrainGVEX in different types of genomic regions. 3' UTR enrichment p value for each cohort relative to REDIportal was determined via Chi-squared test.



Supplementary Figure 2-7. WGCNA of BrainGVEX and CMC cohorts.

a-b, Dendrograms of all WGCNA testable editing sites and their respective modules for the BrainGVEX cohort **(a)** and the CMC cohort (**b**). **c,** Correlation between eigengenes of a WGCNA module and disease condition for the top five largest modules detected in each cohort. Bars are colored by module size. The total number of sites in the largest disease-correlated modules are labeled.

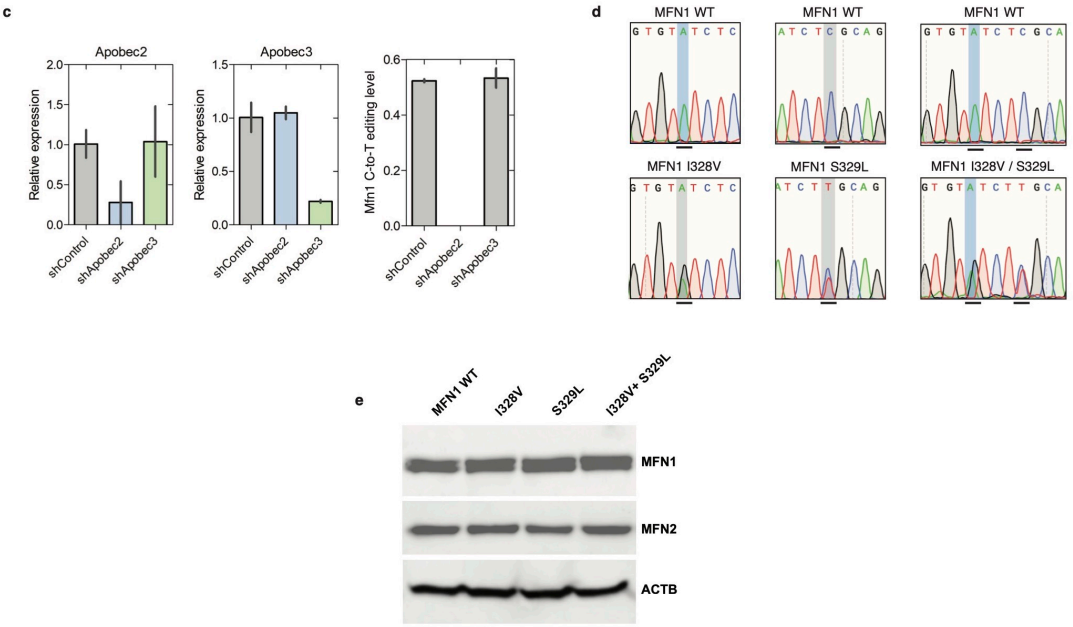
a

Species	A.to.G editing	C.to.T editing	Source
Zebrafish	8.0% - 40%	9.0% - 33%	See Methods, Wong R. & Godwin J. *Genomic Data* (2015)
Mouse	27%	57%	Danecek, P. et al. *Genome Biology* (2012)
Rhesus Monkey	7.5-30%	--	Ye L.Q. et al. *Science Bulletin* (2017)

b

BrainGVEX: Contengency table of total read counts covering A.to.G and C.to.T MFN1 loci		Total Reads with G vs. A		Odds
		A	G	
Total Reads with T vs. C	T	2,251	708	0.3145
	C	4,500	305	0.0678
			Odds Ratio	4.6406
			P value	P < 1.0e-4*

*Exact P-value estimates are unreliable for those below 1.0e-4



Supplementary Figure 2-8. Further investigation of MFN1 C-to-T and A-to-G editing.

a, C-to-T and A-to-G editing levels in zebrafish and mouse. A-to-G editing was examined in Rhesus Monkey. However, C-to-T editing was not studied. Sources are given in the last column of the table. **b,** Number of reads with combinations of A/G vs. C/T nucleotides in BrainGVEX samples. The odds ratio is calculated between the editing-dependent model and editing-independent models to determine editing dependency between the two loci (Chi-squared test was used to determine significance). **c,** Apobec2 and Apobec3 KD conducted in mouse N2a cells. Bar plots show average expression or C-to-T editing levels. Error bars correspond to ±SEM. **d,** Confirmation of prime editing in HEK293T cells of MFN1 via Sanger sequencing of genomic DNA. A-to-G and C-to-T loci are highlighted. **e,** Western blot of MFN1 and MFN2 expression in HEK293T cells. The panel for MFN1 displays two bands capturing MFN1 (bottom) and MFN2 (top). The two proteins, sharing 63% homology, are both captured by the MFN1 antibody.

2.11 SUPPLEMENTARY TABLES

Cohort	Ethnicity	Brain Region	RNA Library type (paired end)	SCZ (after QC)	Control (after QC)	Total (after QC)
BrainGVEX	96% Euro, 1% AA, 3% other*	FC	rRNA depleted (stranded)	65	67	132
CMC	79% Euro, 18% AA, 3% other*	DLPFC	rRNA depleted (unstranded)	65	72	137**
HBCC	100% Euro	DLPFC	rRNA depleted (stranded)	18	23	41
HBCC	100% AA	DLPFC	rRNA depleted (stranded)	48	52	100
LIBD	100% Euro	DLPFC	Poly-A selected (unstranded)	16	20	36
LIBD	100% AA	DLPFC	Poly-A selected (unstranded)	18	26	44

*other: Asian, Hispanic, or non-white
**Only samples < 70 years of age were used in the final analysis (see text for details)
FC: frontal cortex
DLPFC: Dorsolateral Prefrontal Cortex

Supplementary Table 2-1. Summary of datasets analyzed in this study.

Ethnicity, brain region, RNA library type, and total samples are shown per cohort. European and African American samples in the HBCC or LIBD cohorts were analyzed separately.

Supplementary Tables 2-2 to 2-4 are attached.

CHAPTER 3 - The genetic landscape of RNA editing and other aspects of gene expression in schizophrenia

3.1 ABSTRACT

Genome wide association studies (GWAS) have been conducted over the past decades to investigate the underlying genetic origin of neuropsychiatric diseases, such as SCZ. While these studies demonstrated the significance of disease-phenotype associations, there is a pressing need to fully characterize the functional relevance of disease-associated genetic variants. Functional genetic loci can affect transcriptional and post-transcriptional phenotypes such as RNA editing, splicing, and gene expression that may contribute to disease pathology. Here, we investigate the associations between genetic variation and RNA editing, splicing, and gene expression through identification of quantitative trait loci (QTL) in the CommonMind Consortium SCZ cohort. We find that editing QTL (edQTL), splicing QTL (sQTL) and expression QTL (eQTL) possess both unique and common gene targets, which are involved in many disease-relevant pathways, including brain function and immune response. We investigate each QTL-type in both European (EU) and African American (AA) populations, and find that AA-specific QTL are associated with larger effect sizes. In addition, we observe that the mitochondrial RNA binding protein AKAP1 had enriched binding sites among edQTL, including a variant shared by all 3 types of QTL. Finally, we conduct colocalization with various brain disorders and find that all QTL have top colocalizations with SCZ and related neuropsychiatric diseases. This work presents the investigation of multiple QTL types in

parallel and demonstrates how they target both distinct and overlapping SCZ-relevant genes and pathways.

3.2 INTRODUCTION

Schizophrenia (SCZ) is a neuropsychiatric disorder that affects 1% of the population and is characterized by complex behavioral and cognitive symptoms, such as hallucinations and disoriented thinking[141]. The disease is assumed to arise from genetic and/or environmental disruption of brain development. SCZ has profound impact on both the individual and broader society, resulting in billions of dollars per year in economic costs[141,142]. Understanding the underlying mechanisms of SCZ will lead to the development of novel therapies, for which there is a pressing need[142]. With the availability of large case-control cohorts, rapid progresses have been made towards unveiling the genetic basis of the disease[49,143,144] . However, it is challenging to identify specific downstream disease mechanisms for this complex polygenic disorder, with hundreds (or possibly thousands) of distinct genetic loci involved at the population level[141].

While disease-related genome wide association studies (GWAS) uncover relationships between genetic loci and traits of interest, they are often uninformative of the specific functional pathways impacted by complex polygenetic disorders like SCZ. For this reason, association between genetic variants and molecular phenotypes of diseases are increasingly examined. Identification of quantitative trait loci (QTL), such as splicing QTL (sQTL), expression QTL (eQTL), and RNA editing QTL (edQTL), has been conducted for various neuropsychiatric disease cohorts[53,56,76,145]. For example, sQTL were shown to colocalize to SCZ GWAS loci[56], and both sQTL and eQTL identified in

fetal brains associate with Autism (ASD) and SCZ[145]. The latter study also revealed multiple sQTL-associated RNA binding proteins (RBPs), which may regulate spicing by binding to genetic loci of interest[145]. More recently, the importance of edQTL detection in diseases is gaining recognition. For examples, edQTL detected across GTEx tissues showed a strong association with inflammatory diseases[146], and those found in Alzheimer's disease (AD) were predicted to contribute to the differences observed in RNA editing between European (EU) and African American (AA) individuals[76]. In addition, edQTL and eQTL have been identified in SCZ and shown to colocalize to SCZ GWAS loci[53].

While edQTL, sQTL, and eQTL detection in various cohorts has been conducted, the detection of all three types of loci within the same cohort is rare. Indeed, to our knowledge, these three QTL types have not been studied concurrently in any brain disorder. Since gene expression, splicing and RNA editing are regulated by both distinct and inter-related mechanisms[10,147], we hypothesize that investigation of the 3 molecular aspects in the same cohort may identify genetic loci that affect multiple processes and those that are unique to one process. To this end, we analyzed edQTL, sQTL, and eQTL in SCZ from the CommonMind Consortium[83,148] (CMC) for both AA and EU ethnicities. We highlighted QTL of interest that were associated with all three phenotypic categories, which we refer to as "seedQTL", and observed that distinct QTL between ethnicities showed higher effect sizes in AA. Additionally, we identified RBPs with binding sites enriched for QTL, and colocalized QTL with GWAS summary statistics for various neuropsychiatric diseases of interest. This study provides a foundation to a broader

understanding of the relationship between genetic loci and disease-relevant transcriptional and posttranscriptional pathways.

3.3 RESULTS

3.3.1 Quantification of RNA Editing, Splicing, and Expression in SCZ and Controls

We first quantified transcriptome-wide RNA editing, gene expression and splicing in SCZ patients and controls. To this end, we used RNA-seq data of dorsolateral prefrontal cortex (DLPFC) samples in the CMC consortium[148]. As we sought to characterize QTL for both EU and AA ethnicities, we conducted all downstream analyses for the two groups separately. A total of 434 EU and 85 AA samples were obtained and then filtered by our stringent quality control procedures[149]. In our previous work[149], we observed that samples over the age of 70 in CMC displayed lower quality metrics, such as RIN and mRNA proportion. Therefore, we chose to exclude 218 samples over the age of 70 years from the EU cohort. We ensured that metadata variables did not differ between SCZ and control samples for each ethnicity, respectively[149] (Supplementary Fig. 3-1, Supplementary Fig. 3-2). Due to low sample size in the AA cohort, we chose to conduct strict QC for this cohort regardless of age (although 75% of samples were < 70 years-old). Following these procedures, we retained 140 EU and 65 AA high quality samples matched for meta data between SCZ and controls (Fig. 3-1, Supplementary Fig. 3-2).

With the above datasets, we applied our previously established de novo RNA editing detection pipeline[86,127,150] and identified 4,513,844 and 1,607,763 RNA editing sites in the EU and AA cohorts, respectively. For the edQTL analysis, we focused on

editing sites that occur in at least 10% of samples per ethnicity, as QTL detection is most meaningful for loci occurring in multiple samples. These editing loci are referred to as "common sites". An additional editing level variance cutoff of > 0.005 was applied to retain editing sites demonstrating variation across the cohorts. A total of 134,696 and 56,280 common RNA editing sites were used for downstream edQTL detection in the EU and AA cohorts, respectively (Fig. 3-1). The RNA-seq data was further analyzed to calculate gene expression (RPKM) and quantify intron excision ratios (PSI, via LeafCutter[151]) (Methods). We retained 21,772 and 22,220 genes, 109,939 and 75,615 introns in the EU and AA samples, respectively (Methods).

3.3.2 Identification of Quantitative Trait Loci

To carry out QTL analysis, we obtained high-density genotype data for all samples in the CMC cohort from the PsychENCODE Knowledge Portal[83,148]. After normalization and filtering via the PLINK software[152], 5,456,564 and 7,317,154 SNPs were retained in the EU and AA cohorts, respectively (Methods, Fig. 3-1). We conducted PCA on the genotype data for each cohort to determine genotype covariates. Hidden covariates for gene expression, splicing, and RNA editing sites were determined via the hidden covariates with prior (HCP) methodology[153]. We adopted the FastQTL method[154] to identify cis-edQTL, cis-sQTL, and cis-eQTL using known and inferred covariates (Methods, Fig. 3-1). The total numbers of each cis-QTL type and their corresponding targets in both ethnicities are summarized in Table 3-1.

3.3.3 Characterization and comparison of cis-edQTL, cis-sQTL, and cis-eQTL

With the QTL results (Table 3-1, Supplementary Fig. 3-3a, b), we next compared the significance of each QTL type at different distance ranges relative to their corresponding molecular targets. We observed that QTL closer to their targets showed more significant associations with their targeted traits for all three QTL types (Fig. 3-2a-c, Supplementary Fig. 3-3c-e). This observation is consistent with the expectation that cis-acting genetic variants regulating gene expression, splicing or RNA editing tend to reside in the close vicinity of their targets. Additionally, we examined the distances between the QTL and features of their regulatory targets. This analysis showed that, compared to random control SNPs (Methods), edQTL were significantly closer to editing sites (Fig. 3-2d), eQTL were significantly closer to transcriptional start and end sites (Fig. 3-2e), and sQTL were closer to exon start and exon end regions (Fig. 3-2f). Thus, the above results support the validity of the QTL analyses.

Since RNA editing, splicing and total RNA expression are inter-related aspects of gene expression[155,156], we examined whether the three types of QTL and their target genes overlap with each other or whether they are largely distinct. In the EU cohort, we observed that the QTL of all three types significantly overlapped with each other (Fig. 3-2g), although many QTL uniquely target one category. Interestingly, two SNPs were identified to be shared as sQTL, eQTL, and edQTL (s/e/edQTL) (Fig. 3-2g). We hereby refer to such SNPs as seedQTL. For example, one seedQTL (rs146498205) is associated with editing level (exonic editing site), splicing, and expression of the same lncRNA gene, RP11-156P1.3 (Fig. 3-2h, i). Importantly, the expression of this lncRNA has been associated with intracranial volume through TWAS[145] and was found to be a diagnostic

and therapeutic marker in hepatocellular carcinoma[157]. For the AA cohort, no seedQTL were identified (Supplementary Fig. 3-4a). Although few pairwise overlaps were identified between the QTL in AA, they also reached statistical significance, similarly as in the EU cohort (Supplementary Fig. 3-4a).

The three types of QTL analyses demonstrated a larger number of overlaps among target genes than at the level of specific QTL, all of which were statistically significant in the EU cohort (Fig. 3-2j). In addition, 33 genes were targeted by all three QTL categories (Fig. 3-2j). Gene ontology (GO) enrichment analysis of the 33 genes revealed multiple GO categories relevant to SCZ, such as toll-like receptor signaling[158] and nervous system development (Fig. 3-2k). Unlike the EU cohort, we did not see any genes targeted by all three QTL categories in AA. However, we see a significant overlap in target genes between edQTL and sQTL, and between sQTL and eQTL (Supplementary Fig. 3-4b).

Next, we asked whether the target editing levels (EL), PSI, and RPKMs of overlapping QTL (seedQTL, s/eQTL, s/edQTL, or e/edQTL) showed correlations with one another. Interestingly, the aforementioned loci at rs146498205 was the only seedQTL that had significant correlations for all 3 pairs of comparisons: EL vs. PSI, EL vs. RPKM, and PSI vs. RPKM (Methods, Supplementary Fig. 3-4c). The second seedQTL identified in this study only showed association between EL and PSI. Finally, 25 out of 129 e/sQTL showed significant associations between target PSI and RPKM (Supplementary Fig. 3-4c). Thus, our results suggest that there exists overlap between the target genes and the associated genetic loci of the three types of QTL. Nonetheless, it is notable that the vast majority of genetic loci were not shared across different types of QTL, possibly reflecting

the distinct regulatory mechanisms underlying RNA editing, splicing and mRNA abundance.

3.3.4 Distinct and shared QTL in European and African American populations

We next explored whether the same QTL were detected for the EU and AA populations. As shown in Fig. 3-3a-c, all three QTL types had a significant number of overlapping loci between the two ethnicities. Thus, we investigated the directionality of effect that QTL had on their targets for the overlapping and non-overlapping QTL, respectively. To this end, effect size values were utilized to characterize the magnitude and directionality of the association between each QTL allele and phenotype (Methods). We observed that QTL sharing the same targets in AA and EU also showed highly correlated effect sizes on their respective QTL targets (Fig. 3-3d-f). This result strongly supports that genetic factors are the main drivers for the variations of the corresponding molecular phenotypes associated with these QTL.

Intriguingly, for edQTL and sQTL that are unique to AA or EU, we observed overall higher effect sizes in AA compared to EU (Fig. 3-3d-f). For example, there exist more edQTL with positive effect sizes than negative ones, which suggests that alternative alleles often associate with higher editing in the AA cohort (Fig. 3-3d). Similar observations were made for the effect sizes of distinct sQTL and eQTL between the two ethnicities. In general, AA individuals showed higher PSI and RPKM values associated with their alternative alleles. This observation supports the hypothesis that nonoverlapping genetic variants in the ethnicities can show distinct impact on their target phenotypes.

3.3.5 RBP Enrichment in QTL Regions

As post-transcriptional steps of RNA processing, both RNA editing and splicing are closely regulated by a repertoire of cis-acting elements associated with trans-factors (e.g., RBPs). Increasing evidence showed that genetic variants often affect post-transcriptional regulation by altering cis-elements, such as RBP-binding sites[159–161]. Thus, we next examined the enrichment of edQTL and sQTL within the binding sites of RBPs. To this end, we analyzed the binding peaks of 120 and 103 RBPs in the K562 and HepG2 cells, respectively, with eCLIP-seq data generated by the ENCODE consortium[162,163] (Methods). Using random control SNPs as backgrounds, we identified a number of RBPs that preferentially bind to edQTL or sQTL (Fig. 3-4a, b), with the most significant enrichment observed for the A-Kinase anchor protein 1 (AKAP1) associated with edQTL and the DEAD-box helicase 3 X-linked (DDX3X) protein for sQTL.

AKAPs bind and target protein kinase A (PKA) to specific compartments of the cell. Specifically, AKAP1 localizes PKA to the mitochondrial membrane, and is highly involved in mitochondrial oxidative metabolism[164,165]. We identified four edQTL overlapping AKAP1 binding sites that are associated with editing at five different loci (Fig. 3-4c). Interestingly, one of these overlaps is with the seedQTL previously highlighted at rs146498205 (Fig. 3-2h). Thus, AKAP1 binding at this region may influence not only editing, but also expression and splicing of the lncRNA RP11-156P1.3. We observed that both the QTL and the editing site target are in inverted ALU regions, making them highly likely to form a dsRNA structure. To further support this hypothesis, we conducted BLASTn[166] alignment on the ALUs and identified a 59 bp region with 73% identity between

the two inverted ALUs. These portions of the ALU regions contained both the seedQTL and the target editing site (Supplementary Fig. 3-5).

AKAP1 also binds to an edQTL region in the 3'UTR of Cathepsin B (CTSB) (Fig. 3-4c, d), a lysosomal cysteine protease that has been associated with neuropathology in disorders such as Alzheimer's disease, traumatic brain injury, and dementia[167]. Importantly, the edQTL targets two editing sites within the same 3' UTR of CTSB, which have lower editing associated with the QTL alternative allele (Fig. 3-4d, e). Both the editing sites and edQTL are located in ALU elements. As a result, the region encompassing these sites fold into a strong dsRNA structure based on RNAfold[168] (Fig. 3-4f). In the predicted dsRNA structure, the edQTL and one of its editing target sites are counterparts in a loop region flanked by highly base-paired sequences. Furthermore, we observe the lowest minimum free energy (MFE), which yields a more stable dsRNA structure, when we include the reference allele and editing at both target ES compared to no editing (Fig. 3-4f). Upon investigation of all possible structures, we found two major 3'UTR CTSB configurations, determined by the edited or unedited nucleotide in the loop region (Supplementary Fig. 3-6a, b). The stability of these two structures are determined by the pairing of the edQTL and its opposing editing site (Supplementary Fig. 3-6a-h). These examples highlight the possible interplay between RBP binding, edQTL, and their regulation of target editing sites.

Similar to edQTL, sQTL also showed enriched overlap with a number of RBP binding sites, the top one being DDX3X (Fig. 3-4b). DDX3X overlapped with 9 sQTL targeting 20 introns (Supplementary Table 3-1). This RBP is a stress granule protein and has many known nuclear roles including transcriptional regulation, mRNP assembly, pre-

mRNA splicing, and mRNA export[169,170]. Additionally, it has been associated with brain developmental abnormalities and multiple cancers, including brain tumor development in intellectually disabled females[170,171]. Another RBP enriched with sQTL was G3BP1, which has an essential role in stress granule formation, is part of the RAS signal transduction pathway, and promotes innate immune response[172]. Next, we asked whether genes targeted by the RBPs with enriched sQTL (Fig. 3-4b) showed interesting functional pathways. To this end, we conducted GO enrichment of 34 genes whose introns were targeted by the RBP-associated sQTL. We found a number of functional pathways related to immune response and apoptosis, such as natural killer cell lectin-like receptor binding, modulation by virus of host morphology or physiology, and regulation of apoptotic process (Fig. 3-4g).

3.3.6 QTL Colocalization with GWAS of Neurological and Neuropsychiatric Diseases

To further investigate the disease relevance of the QTL, we carried out colocalization analysis with GWAS of neurological and neuropsychiatric diseases. Specifically, we aggregated GWAS summary statistics for SCZ, BPD, major depressive disorder (MDD), ASD, attention deficit hyperactivity disorder (ADHD), post-traumatic stress disorder (PTSD) and alcohol dependency (ALD, as a negative control) from the Psychiatric Genomics Consortium (PGC) (Supplementary Table 3-2)[143,144,173–177]. These GWAS summary statistics were colocalized to QTL in both EU and AA using the COLOC software[178,179] (Methods).

Among the disorders included in this analysis, SCZ, BPD and MDD yielded the most colocalizations in the three QTL types (Fig. 3-5a). This observation is consistent

with the fact that SCZ, BPD, and MDD are known to share specific disease etiologies including similar causal factors and genetic loci linked to the diseases[180,181]. Aggregating colocalizations identified with all disease GWAS and all QTL-trait associations, we identified a total of 8, 33, 53 unique colocalizations for edQTL-ES, sQTL-intron, and eQTL-gene respectively (PP H4 > 50, Methods, Supplementary Table 3-3). These colocalizations were in 6, 19, and 35 unique disease-GWAS loci respectively (within ±1MB of GWAS P < 5e-8, Supplementary Table 3-3). The edQTL colocalization with the highest posterior probability was located at rs34819784 and colocalized with SCZ GWAS (Fig. 3-5b, Supplementary Table 3-3). This SNP in the protein kinase PRKD3 is associated with an editing site in the 3'UTR of CEBPZ-AS1, which is known to be an integral part of the mitochondrial membrane[182]. Although CEBPZ-AS1 has not been well studied, its association to editing and SCZ further support the importance of mitochondrial regulation in neuropsychiatric diseases[149].

We also identified an edQTL that colocalized to SCZ GWAS at rs1058298 located in the adenosine deaminase like (ADAL) gene (Supplementary Fig. 3-7a). ADAL protein catalyzes the hydrolysis of the free cytosolic methylated adenosine nucleotide N(6)-methyl-AMP (N6-mAMP) to produce inositol monophosphate, making it essential for m6a regulation[183]. In addition, we observed two edQTL that colocalize to MDD GWAS at rs56238203 and rs73168402 located in the gene LSAMP, a possible gene target for neuropsychiatric diseases that has been associated to panic disorder and social behavior[184–186]. These genetic loci target two editing sites in the lncRNA RP11-384F7.2 (Supplementary Fig. 3-7b, c). While the lncRNA has not been well studied in the brain, it is a known risk indicator for lung cancer[187].

Among the colocalizations between GWAS diseases and eQTL and sQTL, we identified many target genes with interesting functionalities (Supplementary Table 3-3). Examples of top colocalizations for the sQTL and eQTL are shown in Figs. 3-5c and 3-5d (PP H4 > 90, GWAS P < 5e-8) between SCZ and the target genes PSMA4 and FURIN, respectively. While the PSMA4 gene has been linked to SCZ, tobacco addiction, and cancer[188,189], FURIN is known to contain SCZ risk variants that synergistically affect synaptic function[190]. Finally, we highlight a strong colocalization between SCZ GWAS summary statistics and the eQTL at rs6439649 affecting expression of the PCCB gene, which encodes the beta subunit of the mitochondrial enzyme and for which expression difference has been observed in both SCZ and ASD (Supplementary Fig. 3-7d)[191].

Due to low AA sample sizes in disease GWAS and brain RNA-seq, we did not observe significant colocalizations between QTL and SCZ, BPD, or MDD in this ethnicity. The lack of results indicates the necessity for more comprehensive GWAS in the AA population. Overall, the above colocalization results demonstrate the potential disease relevance of the three types of QTL, with top hits occurring in genes with functions related to the brain and mitochondria.

3.4 DISCUSSION

We present the first study, to our best knowledge, of edQTL, eQTL, and sQTL within the same SCZ cohort and for two different ethnicities. We detected largely distinct QTL of each type, but with statistically significant pairwise overlaps between the 3 QTL types. Upon comparison between QTL in AA and EU, we observed common QTL of each type between the two ethnicities, for which we also observed similar effect sizes. In contrast,

QTL unique to one ethnicity showed higher effect sizes in AA. Finally, we identified RBPs enriched with edQTL and sQTL, and observed colocalizations at various loci between all QTL types and SCZ, BPD, and MDD.

Our concurrent examination of edQTL, sQTL and eQTL detected significant pairwise overlaps between the 3 QTL types. We also identified a small number of seedQTL associated with all 3 QTL categories. These observations suggest that genetic regulations of RNA editing, splicing and RNA abundance may be inter-related, sharing common mechanisms. The example seedQTL at rs146498205 and its target are located in inverse ALU elements that may form a dsRNA structure (Fig. 3-2h). We postulate that structural alteration by the QTL may explain its association with all 3 molecular phenotypes, although future work is needed to confirm this possibility. Apart from individual loci overlap, the target genes of the 3 types of QTL overlap significantly, enriched in pathways related to brain development and immunity. This observation supports the hypothesis that genetically driven variations in RNA editing, splicing and abundance influence similar functional pathways. Nevertheless, existence of a large number of QTL unique to each molecular trait indicates that the 3 molecular processes make diverse contributions and represent distinct (but inter-related) layers of gene regulation. Therefore, our results support the need to investigate multiple QTL types in parallel.

In our previous work[149], we observed distinct patterns in RNA editing of SCZ patients between the AA and EU populations. Specifically, SCZ patients in the AA cohort had enhanced RNA editing levels compared to controls, which was opposite to the reduced editing observed in SCZ patients of the EU cohort. We hypothesized that

differences in the general genetic background between the two ethnicities may have led to this distinction. Our current analysis showed that, although the EU and AA cohorts had significant overlap of all QTL, the many unique QTL between the two ethnicities support existence of ancestry-specific regulation of the three molecular phenotypes. Furthermore, we observed a higher effect size for QTL unique to AA subjects compared to those unique to EU subjects (Fig. 3-3d-f). A similar trend was reported for AA-specific eQTL in previous literature[192,193]. While the specific reasoning behind these ancestry-specific observations are elusive, it has been demonstrated that the AA genomes possess greater genetic diversity, population differentiation, and difference in LD structure when compared to those of EU[192]. Therefore, genetic ancestry is a relevant factor for most QTL-phenotype analyses, and may influence downstream transcriptional and post-transcriptional regulation. It should be noted that we observed much smaller numbers of QTL for the AA cohort than the EU cohort, likely due to the smaller sample sizes for AA. This limitation calls for enhanced data acquisition efforts across multiple ethnicities, which will greatly facilitate future large-scale integrative analyses.

The identification of QTLs further enabled investigations of their relevant regulatory mechanisms. To this end, we focused on the possible involvement of RBPs in edQTL and sQTL regulation. We highlighted one RBP, AKAP1, binding to the seedQTL at rs146498205 and a few other edQTL. AKAP1 has many essential functions related to neuronal regulation and mitochondrial function. For example, inhibition of mitochondrial fission by an outer mitochondrial complex containing this RBP protects neurons from ischemic stroke through maintenance of respiratory chain activity, inhibition of superoxide production, and delay of Ca^{2+} deregulation[194]. RBPs enriched at sQTL included stress

granule proteins DDX3X and G3BP1, which are both associated with neurodevelopmental disorders, including ASD, intellectual disability, and ADHD[195,196]. The gene targets of these and other RBPs enriched at sQTL were involved in pathways related to immune response and apoptosis, both relevant to the disease mechanisms of SCZ[197,198]. The above results suggest that edQTL and sQTL are enriched with functional genetic variants that are implicated in disease-related pathways. However, we note that as an association analysis, QTL does not necessarily capture the functional variants that cause alterations in the molecular phenotypes under study. Future methods that pinpoint the causal variants are needed to further understand the genetic regulation of splicing and RNA editing[161].

The colocalization analysis of QTL with GWAS of neurological and neuropsychiatric diseases uncovered a number of QTL with potential disease relevance. As highlighted in the Results, a number of genes related to brain function harbor QTL colocalized with GWAS loci. These findings are consistent with the conjecture that RNA editing, splicing and RNA abundance are all important contributors to SCZ mechanisms, likely explaining a substantial amount of GWAS signals. A previous study identified 11 edQTL in the CMC cohort that colocalized to 6 SCZ GWAS loci[53]. These loci did not overlap with the GWAS-colocalized edQTL we identified. The lack of overlap could be due to the different methods used to identify RNA editing sites, and the relatively small sample size used in our analysis as we excluded samples over 70 years of age in the EU population due to their limited data quality.

In summary, our study highlights the disease relevance of RNA editing, splicing and RNA abundance regulation, making unique and common contributions in different

ethnicities. Indeed, adding further QTL characterization, such as epigenetic QTL, and increasing sample size for EU and AA high quality brain tissue RNA-seq, can bolster such findings. This work motivates future large-scale multi-QTL investigations to more comprehensively reveal the possible interplay between genetic and disease-relevant mechanisms.

3.5 METHODS

3.5.1 Data extraction and Quality Control

Data extraction and processing was conducted as described in our previous work[149]. Briefly, unstranded rRNA-depeleted RNA-seq data from the DLPFC brain region were extracted from the CMC consortium[83,148]. We followed strict quality control procedures to remove sample outliers in RIN, PMI, age, and other biological and technical variables[41]. The quality control procedure was conducted for EU and AA samples separately. Retained SCZ and control groups did not differ significantly in any biological or technical covariates (Supplementary Figs. 3-1, 3-2).

3.5.2 RNA editing detection and normalization

We used our previously developed methods for RNA detection and processing[86,127,150]. RNA-seq was aligned using hisat2 v.2.0.4[126] with default parameters, but allowing no discordant reads. Unmapped reads were extracted for re-alignment of hyperedited reads as described in previous work[41,87]. Briefly, hyperedited reads contain copious mismatches relative to the reference genome, which may prevent them from aligning to the correct genomic position. To address this challenge, all adenosines in both

the reference genome and reads were converted to guanosines. A second round of mapping with hisat2 was completed using the converted genome and reads. Subsequently, all guanosines were changed back to their original adenosines, and uniquely mapped reads were combined with those identified in the first step alignment for downstream processing. After identifying mismatches in the reads relative to the reference genome, several filters were applied to remove loci likely resulting from spurious read mapping or sequencing errors[127]. The final editing sites were required to be supported by at least five samples, and, in each sample, have at least two edited reads and five total reads. The loci occurring in at least 10% of samples within a cohort were labeled as "common" RNA editing sites. Common RNA editing sites in autosomes with > 0.005 variance were retained for downstream analysis. The missMDA R package was used to impute missing data via the principal component methods[199].

RNA editing values were normalized per site as z-scores from a rank normalized distribution. This normalization technique was recommended by MatrixEQTL[200] to remove systematic inflation of test statistics. Specifically, for an editing site *(e)* all samples *(i*) were ranked by editing level (the average was taken for tied rankings). For each site *(e),* average and standard deviation of ranks was used to create a normal distribution. Finally, a per sample *(i)* z-score was calculated according to each site's normal distribution and deemed the normalized RNA editing level for site *(e)* and sample *(i).*

3.5.3 PSI and RPKM Quantification

The Leafcutter software[151] (v 0.2.7) was used to detect reads that span introns and to quantify clusters of spliced introns. Bam files from hisat2 alignment were converted into

junction files, and intron clustering was conducted using default settings in Leafcutter (50 reads per cluster and intron length ≥ 500kb). The prepare_phenotype_table script was then implemented to calculate intron excision ratios and filter out introns used in less than 40% of individuals with < 0.005 variation (default parameters). Intron excision ratios (PSI) were standardized and quantile normalized. We obtained RPKM values from our previous study of RNA editing in SCZ[149]. Briefly, RNA-seq bam files were used to calculate read counts per gene with the HTSeq software (v 0.6.1)[134] and total mapped reads from hisat2. Rank normalization of RPKM values was conducted similarly as for RNA editing values detailed above (see RNA editing detection and normalization). Missing PSI and RPKM values were imputed using missMDA[199], similarly as for RNA editing ratios.

3.5.4 Genotype filtering

Whole-genome genotype data for EU and AA DLPFC samples with standard imputation protocols were extracted from the CMC consortium[148]. The PLINK software (v1.90b6.24)[152] was used to identify sites of interest with the following criteria: 1) bi-allelic SNPs, 2) Minor allele frequency (MAF) > 0.05, 3) Imputation score > 0.8, 4) Hardy Weinberg Equilibrium (HWE) P-value < 1.0e-6, and 5) Missingness < 0.20.

3.5.5 Covariate selection

To measure hidden batch effects and confounders for phenotype data (EL, PSI, and RPKM), we performed covariate detection using the HCP method[153]. This algorithm predicts hidden confounding factors in datasets with known measured covariates. We

calculated HCPs separately for each QTL type and ethnicity. RNA-seq PICARD summary metrics were used as covariates for each cohort, which included median 5' to 3' bias, GC dropout, AT dropout, proportion HQ aligned bases, proportion mRNA bases, and proportion intergenic bases. We calculated 0-50 HCPs in increments of 10 and identified the component number that maximizes the total number of QTL-phenotype associations detected by the FastQTL[154] software (see cis-QTL detection). For the range of HCPs resulting in the maximum associations, we tested increments of 2 and 5 to further pinpoint the most fitting HCP component number (Supplementary Fig. 3-8a-f). We ran PCA on SNPs for the EU and AA cohorts respectively to obtain the top 1-10 genotype PC eigengenes. We included 3 genotype PCs in the final model by testing for population inflation of trans-QTL identified through MatrixEQTL[200] (see cis-QTL detection, Supplementary Fig. 3-8a-f).

3.5.6 cis-QTL detection

The FastQTL[154] software (v2.184) was used to perform cis-edQTL, cis-sQTL, and cis-eQTL mapping in autosomes. A window of $\pm$100 KB from target editing sites was defined as *cis* for edQTL detection, while $\pm$1 MB from the start of genes or introns was tested for eQTL and sQTL, respectively. We included the HCPs (see above, Table 3-1, Supplementary Fig. 3-8a-f), age, RIN, sex, and 3 genotype PCs as covariates in FastQTL's permutation pass mode (1000 permutations) to identify the beta-approximated p-values. The number of genotype PCs was chosen based on the control analysis using trans-eQTL as described in the paragraph below. The output QTL consist of SNPs with the best nominal association per phenotype, yielding p-values through beta

approximation that utilizes a permutation approach to correct for the multiple variants tested per phenotype[154]. We conducted Storey and Tibshirani FDR correction[201] on beta-approximated p-values to account for multiple testing across all phenotypes. Significant edQTL, sQTL, and eQTL were defined as those with FDR q-value < 0.1. QTL-phenotype effect sizes were calculated using Spearman correlation ρ values.

In order to assess possible inflation of p-values, we randomly selected 10 editing sites, introns, or genes and carried out trans-edQTL, trans-sQTL, and trans-eQTL detection using the Matrix_eQTL_main() function (part of the MatrixEQTL (v2.3) R package) with default parameters[200]. All SNPs in the genome were included for the QTL analyses, and the same covariates were used as for the cis-QTL analyses. QQ plots of trans-QTL detection showed zero to very low inflation, indicating that the resulted QTL were not confounded by population stratification (Supplementary Fig. 3-8a-f). The MatrixEQTL detection was carried out for varying numbers of genotype PCs (1-10). We observed the least population inflation via inclusion of 3 genotype PCs. Therefore, 3 genotype PCs were used in the final model for all QTL types and both ethnicities.

3.5.7 Distance to regulatory elements

QTL distance to editing sites, transcriptional start sites (TSS), transcriptional end sites (TES), exon start, and exon end was calculated and compared to random SNPs for the EU cohort. Specifically, all common RNA editing sites detected in the CMC EU samples were considered to determine the closest distance between edQTL and editing sites. TSS, TES, exon start, and exon end sites were obtained from ENSEMBL 99[130]. For genes with multiple isoforms, the most upstream start and downstream end were

documented as the TSS and TES, respectively. Exon start and end sites were first and last coordinates of all aggregated exons documented in ENSEMBL[130]. As controls, random bi-allelic (non-QTL) SNPs of the CMC cohort were used.

For eQTL, edQTL, and sQTL, we extracted 100 sets of random SNPs, each set with the same number of QTL as the test set, respectively. For all QTL and random SNPs, we obtained their distances to editing sites, TES, TSS, exon start, and exon end regions. We calculated an AUC value using the eCDF for distances for each set of random SNPs (and the test set QTL). A Gaussian distribution was generated using the mean and standard deviation of the 100 AUCs from randomizations. The AUC for the test set distances was compared to the Gaussian distribution to calculate enrichment p-values.

3.5.8 Gene ontology (GO) enrichment

GO enrichment was conducted similarly as described in our previous work[149] with the exception that genes matched for expression were determined via average expression calculated from the CMC EU samples. GO enrichment conducted on 34 genes targeted by sQTLs that were enriched with RBP binding sites had an extra step to prune parent GO terms if a child term was significant. This step was implemented to avoid presenting repetitive pathways.

3.5.9 Target correlation between overlapping QTL

We tested the categorical correlation between target EL and PSI, EL and RPKM, and PSI and RPKM for s/eQTL, s/edQTL, e/edQTL and seedQTL in EU (Supplementary Fig. 3-4c). We separated samples into groups with either low or high EL, PSI, or RPKM

to populate a 2x2 contingency table. Low and high normalized editing level cutoffs were chosen to be < -1 and > 0, respectively. Low and high cutoffs were chosen to be < 0.2 and > 0.3 for PSI and > -1 and < 1 for normalized RPKM, respectively. Each contingency table represents a pairwise comparison between two QTL types with the number of samples in each (low/high) category. Pearson's Chi-squared tests were conducted to determine the significance between two target molecular phenotypes.

3.5.10 RBP enrichment

We obtained ENCODE eCLIP data from 120 and 103 RBPs from the K562 and HepG2 cells lines, respectively[163]. eCLIP peaks with fold change (FC) > 3.0, relative to the paired size-matched background input within each region, were combined between both cell lines. We conducted RBP enrichment for edQTL and sQTL separately. For each RBP, we calculated the number of edQTL or sQTL overlapping the eCLIP peaks of the RBP. As a control, we obtained a set of random SNPs from SNPsnap[202], matching the set of edQTL or sQTL in total SNP count, ±5% MAF, ±50% distance to nearest gene, ±50% gene density, and ±50% total 'LD buddies' (SNPs in LD, $r^2 = 0.5$). We then created a 2x2 contingency table for each RBP populated by the total non-overlapping and overlapping counts of the QTL of interest or those of random control SNPs. Fisher's Exact p-values and odds ratios were obtained to determine enrichment significance for each RBP.

3.5.11 Colocalization Analysis

GWAS summary statistics were obtained for SCZ, BPD, MDD, PTSD, Alcohol dependence, ASD, and ADHD from the psychiatric genomics consortium (PGC) (Supplementary Table 3-2)[143,144,173–176,203]. For significant loci ($p < 5e\text{-}8$) in the GWAS summary statistics, we extracted all variants within $\pm$1MB and determined if such variants were edQTL, sQTL, or eQTL (referred to as a query QTL). If confirmed, this $\pm$1MB region was defined as a testable region for colocalization between the specific query QTL's phenotype and GWAS disease. For all testable regions, we conducted extended FastQTL analysis to obtain nominal p-values for associations between the phenotype of interest and SNPs within $\pm$ 1 MB of the query QTL. Nominal FastQTL p-values and GWAS summary statistic p-values for matching SNPs (MAF > 0.05) within $\pm$1 MB of the query QTL loci were input into the COLOC software (v 5.1.0) to test for disease colocalization[179].

The COLOC algorithm obtained posterior probabilities of five hypotheses: H0) no GWAS or QTL signal, H1) GWAS signal only, H2) QTL signal only, H3) GWAS and QTL signal are present but not colocalized, H4) GWAS and QTL signals are present and colocalized. We retain colocalizations with PP H4 > 50, similarly as in previous literature[53], and report all colocalizations in Supplementary Table 3-3. We then obtained colocalization enrichment per QTL and disease, by calculating the percentage of QTL colocalized with disease GWAS among all QTL of the same type (Fig. 3-5a). Colocalizations of interests were visualized using the locuscomparer R tool[204]. Colocalization analysis was run on both EU and AA cohorts separately. However, very few significant colocalizations were identified for the AA cohort. Thus, AA colocalizations are not presented here.

3.6 ACKNOWLEDGEMENTS

We would like to thank Dr. Michael Gandal and Cindy Wen for providing guidance on cis-QTL detection. This work was supported in part by grants from the National Institutes of Health (R01MH123177, R01AG075206 to X.X.). M.C. was supported by the NIH T32LM012424.

3.7 AUTHOR CONTRIBUTIONS

M.C. and X.X. designed the study. M.C. conducted all analyses. X.X. provided supervisory inputs and was the principal investigator.

3.8 COMPETING INTERESTS

The authors declare no competing interests.

3.9 FIGURES

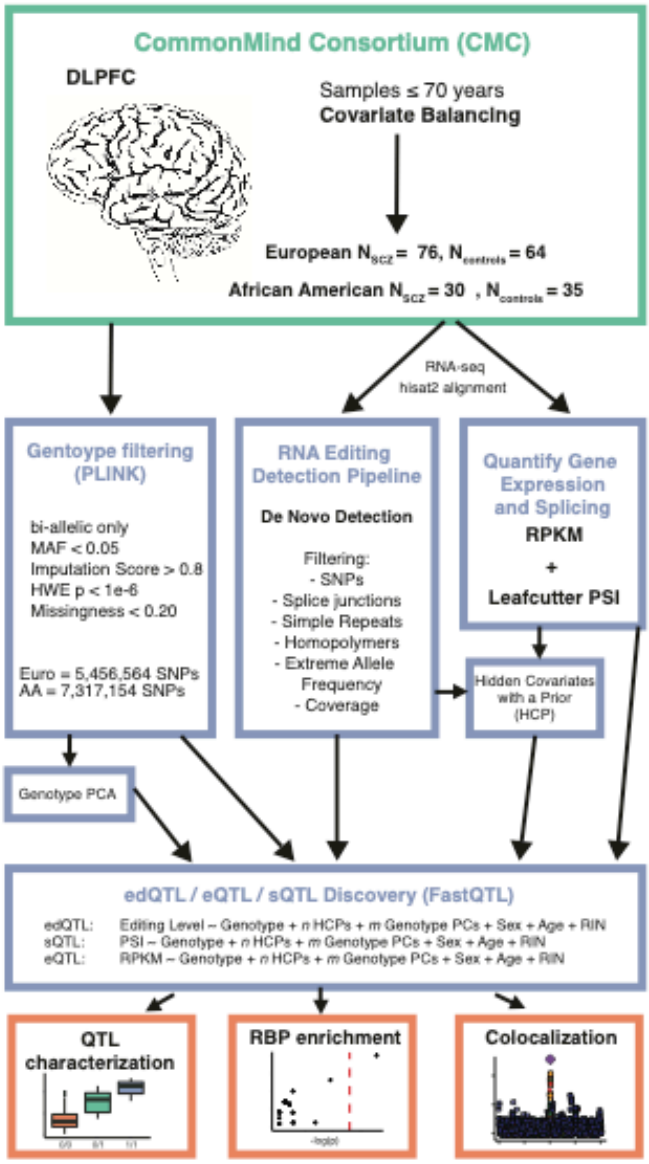


Figure 3-1. Overview of QTL detection and downstream analyses.

DLPFC RNA-seq data was extracted from the CMC consortium and processed for quantification of RNA editing, gene expression, and intron excision values (PSI). Filtered genotypes, hidden covariates, known covariates, and phenotype data were input into linear regression models in the FastQTL software for QTL detection. Finally, downstream analyses were conducted to characterize QTLs, including investigation of relevant regulatory RBPs and disease association through GWAS colocalization.

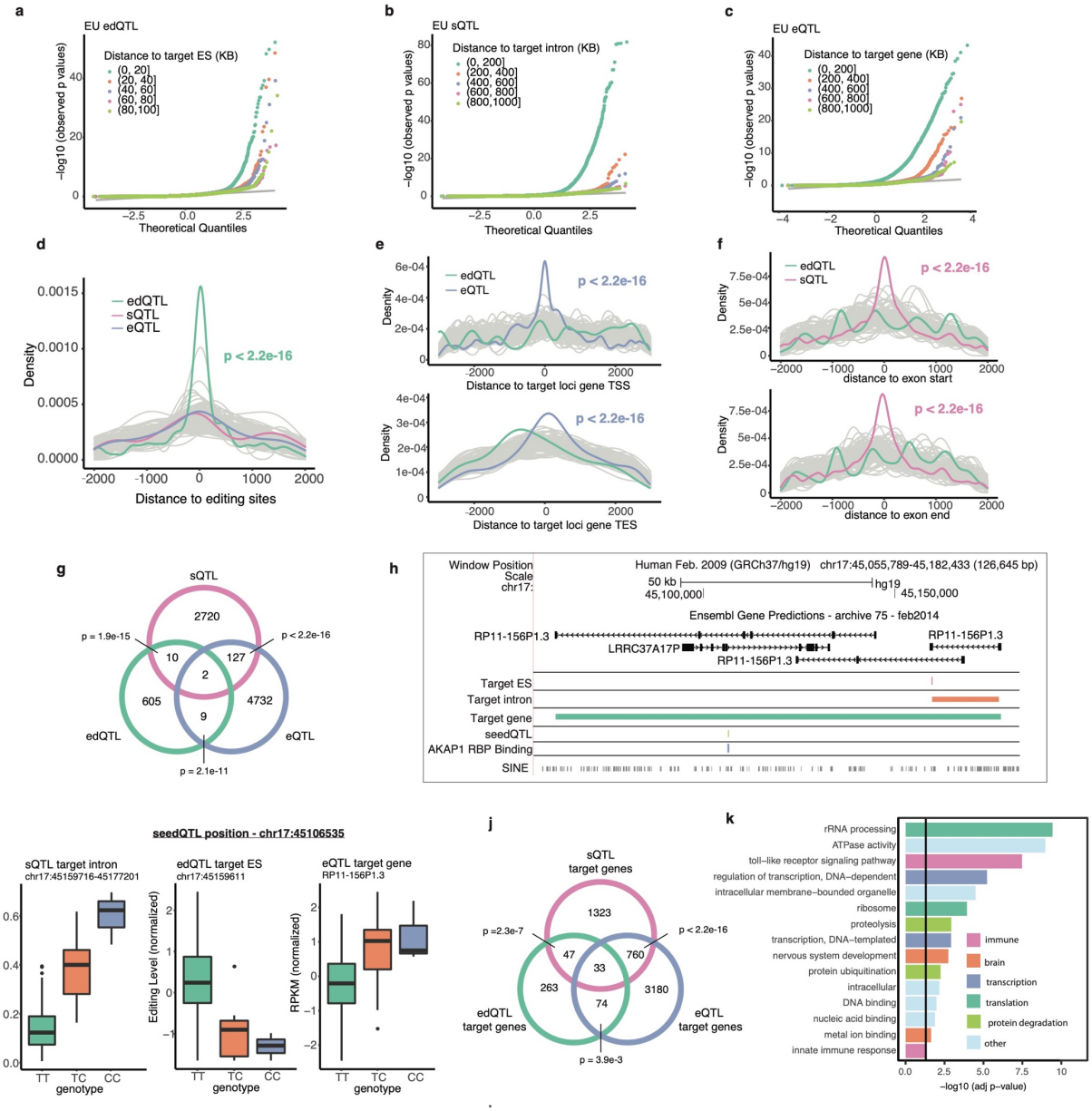


Figure 3-2. QTL characterization.

a-c, QQ plot of edQTL, sQTL, and eQTL detected in EU, grouped by their distance from the associated molecular targets. **d,** Distances between edQTL, sQTL, or eQTL and common editing sites. Grey curves represent the distances between random control SNPs and common editing sites, based on which the p values were calculated (Methods). For visualization, distances of variants within ±2KB from ES are displayed. **e,** Similar to (d), but for the distances of eQTL and edQTL to transcription start sites and transcription end sites. **f,** Similar to (d), but for the distances of sQTL and edQTL to exon starts and exon ends. **g,** Overlap between edQTL, eQTL, and sQTL detected in the EU cohort. Hypergeometric p values are shown to evaluate the overlap between each pair. **h,** Genomic region harboring the

seedQTL rs146498205, its target editing site, target intron region, and target gene region. The AKAP1 RBP binding site is also shown. **i**, Target intron PSI, editing level (EL) of the target editing site (ES), and RPKM of the target gene of the seedQTL rs146498205 for different genotype groups. **j,** Overlap between the genes targeted by the eQTL, edQTL, and sQTL, respectively in the EU cohort. Hypergeometric p values are shown to evaluate the overlap between each pair. **k,** GO enrichment results of the 33 genes targeted by all three QTL types.

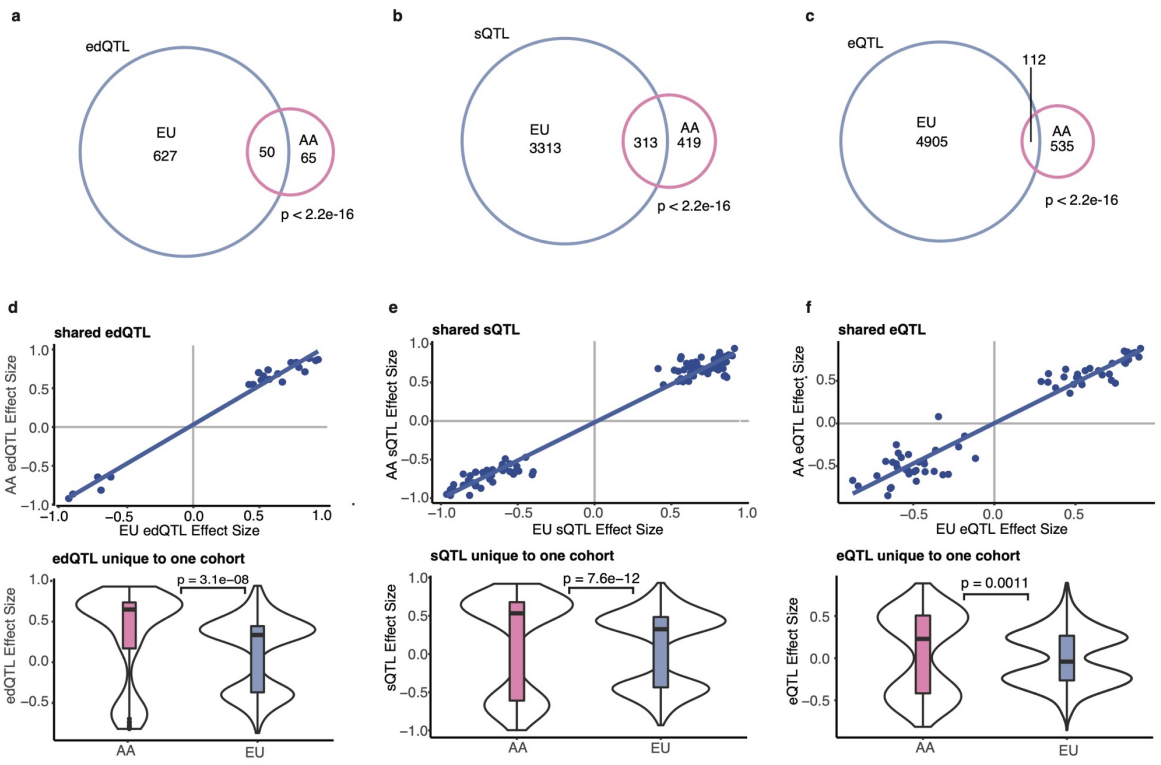
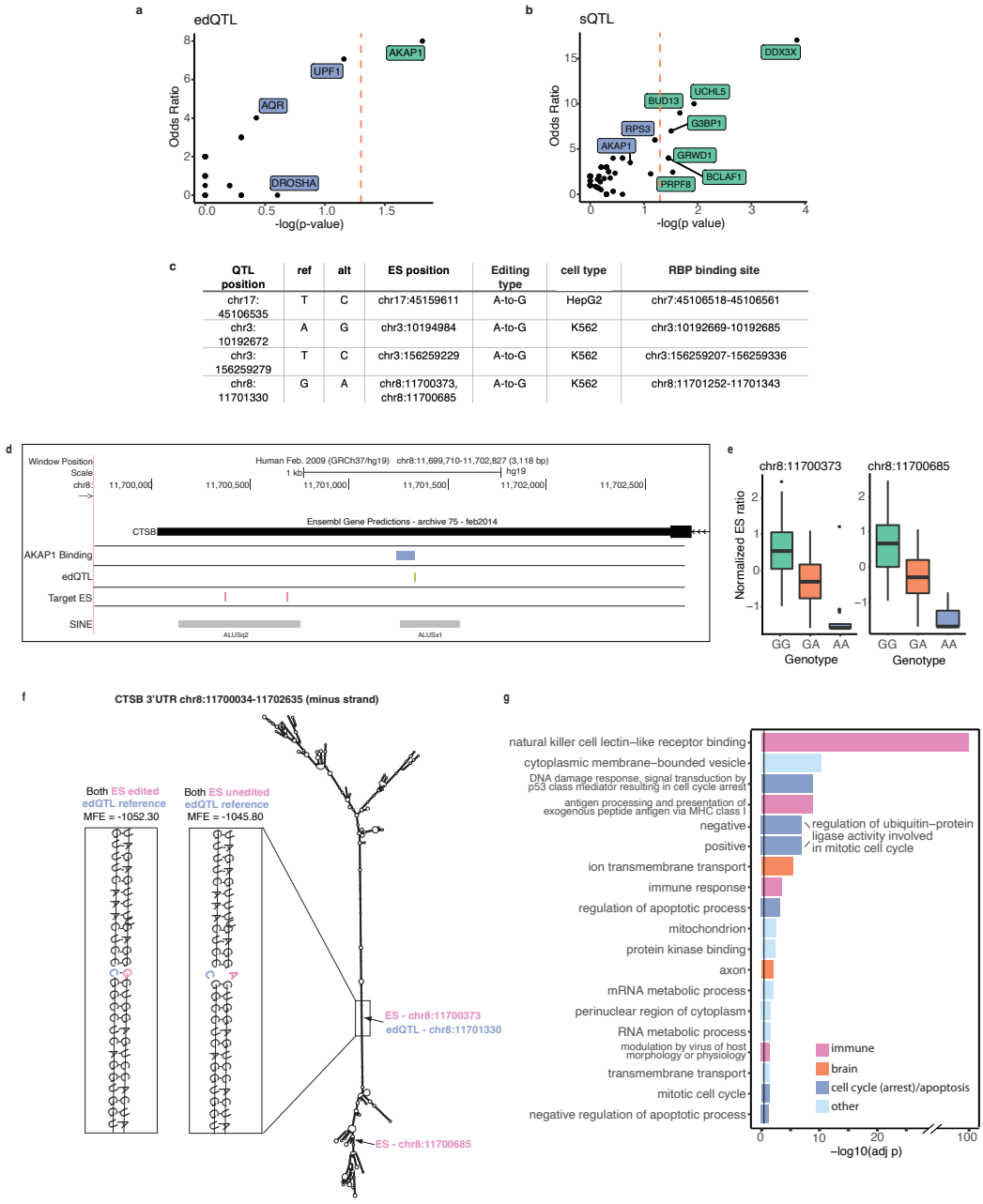


Figure 3-3. QTL detection in EU vs. AA.

a, edQTL overlap between EU and AA . Hypergeometric p value is shown to assess the significance of overlap between ethnicities. **b,** Similar to (a), but for sQTL overlap between ethnicities. **c,** Similar to (a), but for eQTL overlap between ethnicities. **d,** Effect sizes of overlapping edQTL between EU and AA (top). Effect sizes of edQTL unique to AA or EU (bottom). Wilcoxon rank-sum test p values are shown to evaluate difference in effect sizes of unique edQTL between ethnicities. **e,** Similar to (d), but for sQTL. **f,** Similar to (d), but for eQTL.



QTL position	ref	alt	ES position	Editing type	cell type	RBP binding site
chr17: 45106535	T	C	chr17:45159611	A-to-G	HepG2	chr7:45106518-45106561
chr3: 10192672	A	G	chr3:10194984	A-to-G	K562	chr3:10192669-10192685
chr3: 156259279	T	C	chr3:156259229	A-to-G	K562	chr3:156259207-156259336
chr8: 11701330	G	A	chr8:11700373, chr8:11700685	A-to-G	K562	chr8:11701252-11701343

Figure 3-4. Enrichment of edQTL and sQTL in RBP binding sites.

a, Enrichment of edQTL within RBP binding sites. Fisher's Exact test p value and odds ratio for each RBP were calculated via comparisons between edQTL and random SNPs (Methods). **b,** Similar to (a), but for sQTL. **c,** edQTL within AKAP1 binding sites and their target editing sites (ES). **d,** The 3'UTR of CTSB with an AKAP1 binding site, edQTL, and target editing sites. **e,** Normalized editing level of two editing sites associated with the CTSB edQTL, samples grouped by the genotype of the edQTL. RNA editing levels were normalized per site by calculating z-scores from a rank normalized distribution across samples (see Methods). **f,** RNA structure of the CTSB 3'UTR predicted by

RNAfold, given the reference allele of the edQTL and the edited or unedited base at the editing site (ES). The two example structures show the strongest and weakest dsRNA structures according to minimum free energy (MFE). **g,** Enriched GO categories of the 34 genes targeted by sQTLs located in RBPs labeled in (b).

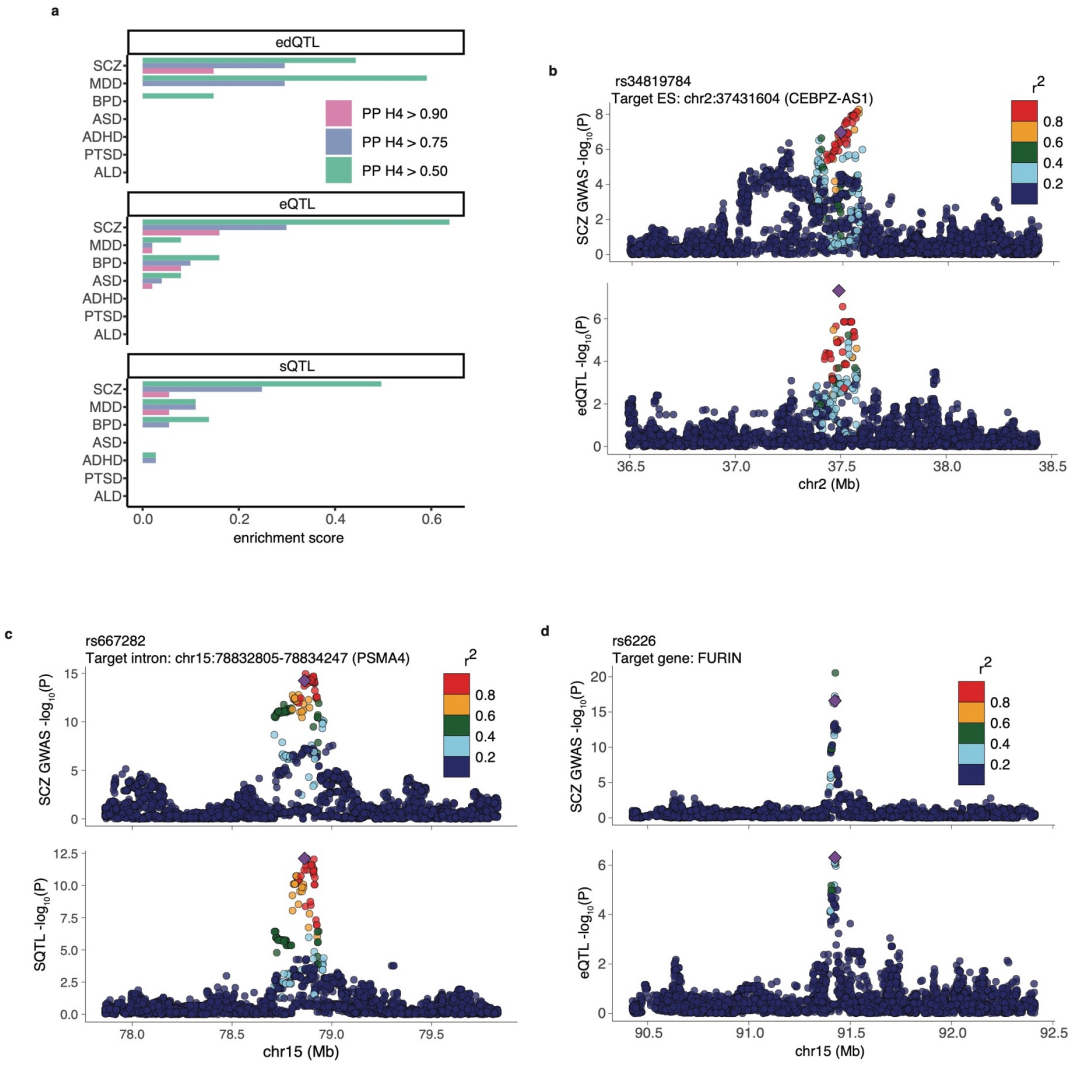
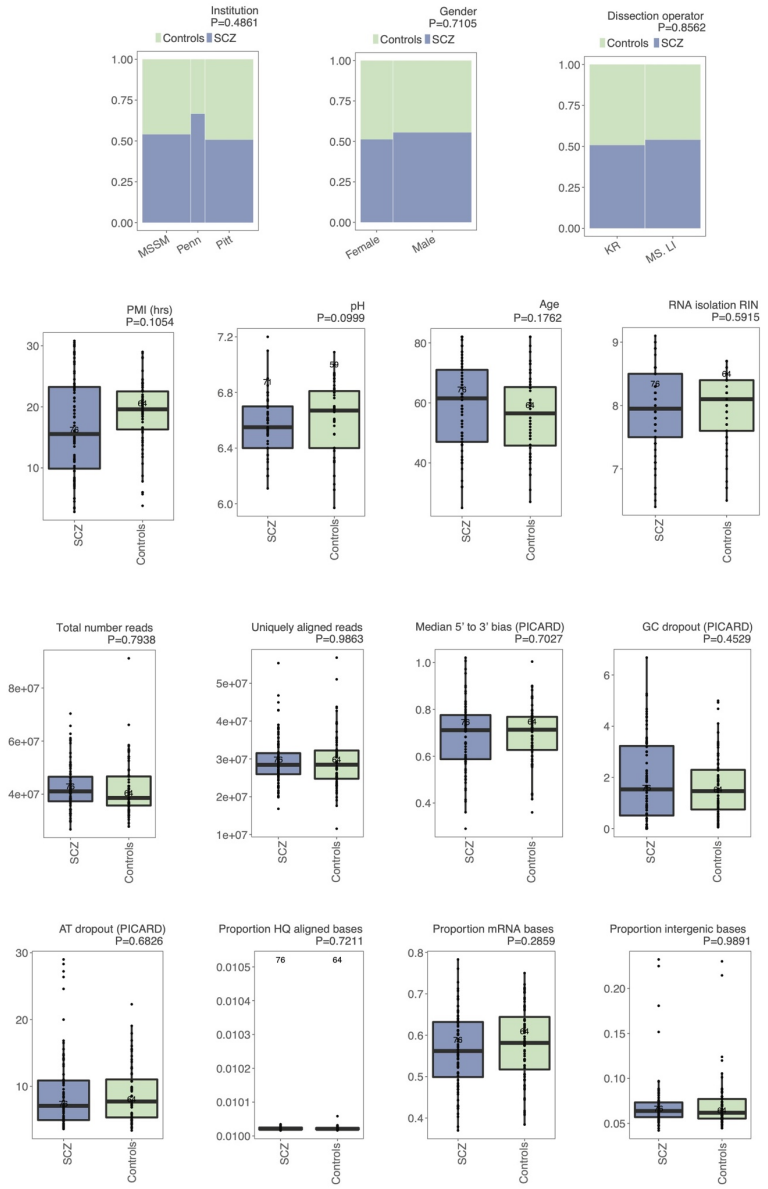


Figure 3-5. QTL Colocalization with GWAS loci of different diseases.

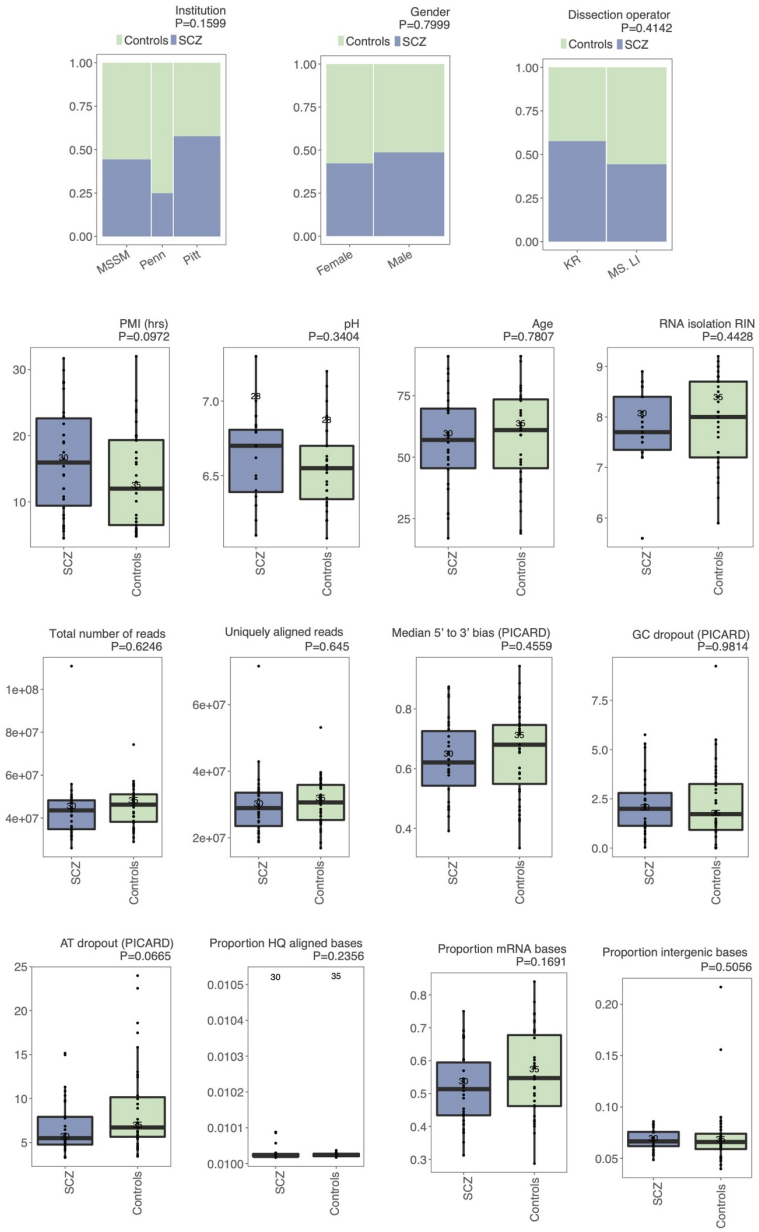
a, Colocalizations detected with GWAS summary statistics for each QTL category. Enrichment score: a ratio between the number of QTL that colocalized to a disease and the total number of QTL of the respective type. **b,** Example colocalization between SCZ GWAS (top) and an edQTL (bottom). The target editing site (ES) is located in the mitochondrial membrane gene CEBPZ-AS1 (PP H4 = 92.5). **c,** Example colocalization between SCZ GWAS (top) and a sQTL (bottom). The target splicing event is located in the gene PSMA4 (PP H4 = 96). **d,** Example colocalization between SCZ GWAS (top) and an eQTL (bottom) of the FURIN gene (PPH4 = 97).

3.10 SUPPLEMENTARY FIGURES



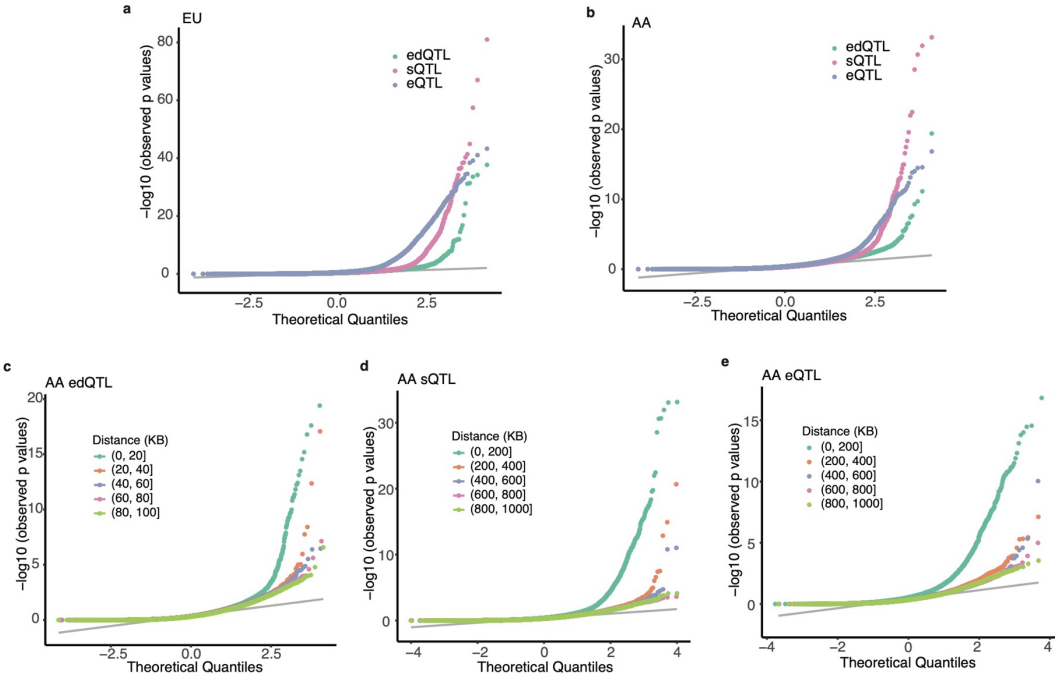
Supplementary Figure 3-1. Comparison of potential confounding variables between SCZ and controls of the EU cohort.

Biological covariates, technical covariates, and RNA-seq mapping metrics for SCZ and control EU samples after completing the sample QC procedure (Methods). P values were calculated via Pearson correlation and Fisher's exact test for numeric and categorical covariates, respectively.



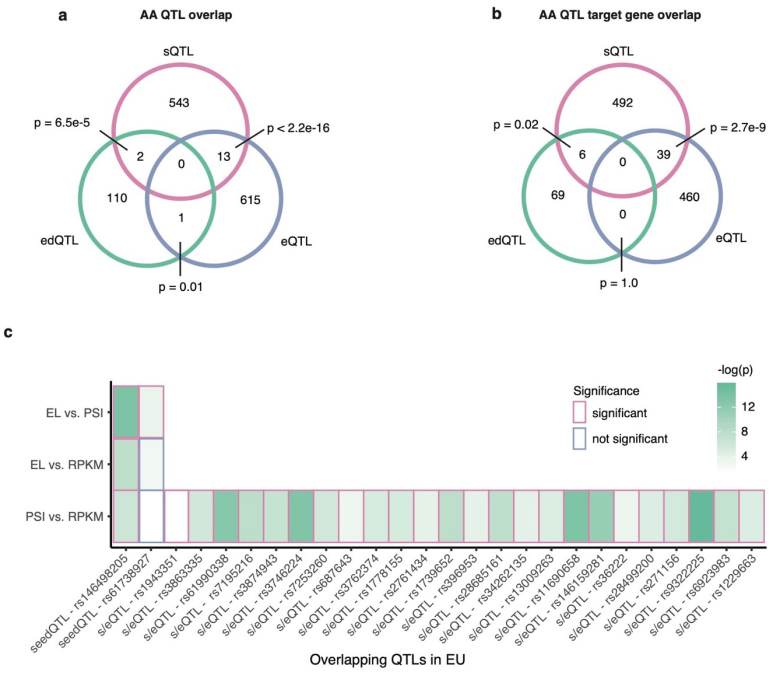
Supplementary Figure 3-2. Comparison of potential confounding variables between SCZ and controls of the AA cohort.

Biological covariates, technical covariates, and RNA-seq mapping metrics for SCZ and control AA samples after completing the sample QC procedure (Methods). P values were calculated via Pearson correlation and Fisher's exact test for numeric and categorical covariates, respectively.



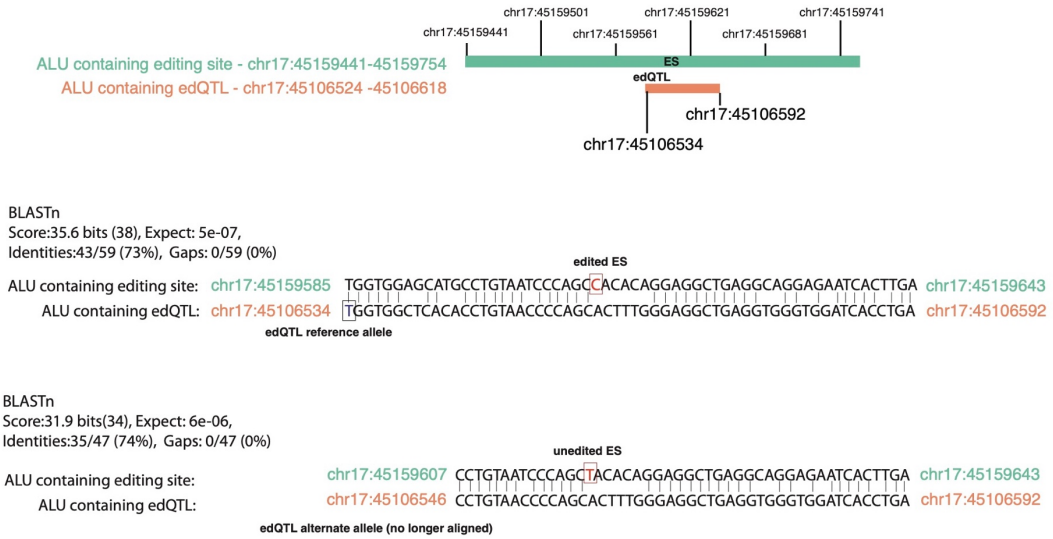
Supplementary Figure 3-3. Supplementary Figure S3. QTL significance in EU and AA cohorts.

a, QQ plot of all QTL detected in EU, grouped by QTL type. **b,** QQ plot of all QTL detected in AA, grouped by QTL type. **c,** QQ plot of edQTL in AA, grouped by their distance to the associated molecular targets. **d**, Similar to (c), but for sQTL in AA. **e**, Similar to (c), but for eQTL in AA.



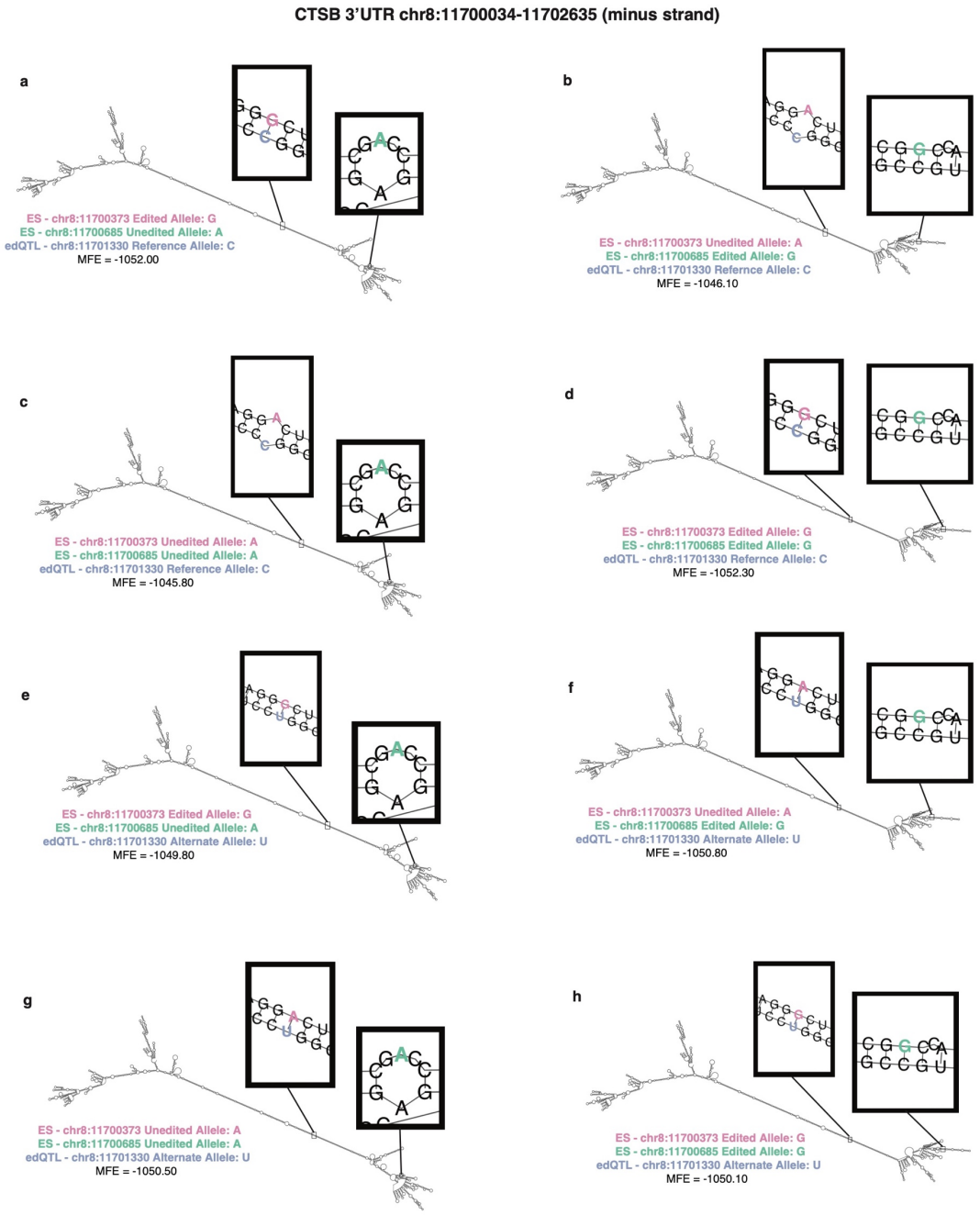
Supplementary Figure 3-4. **Overlaps among the 3 types of QTL or their respective molecular targets.**

a, Overlaps among eQTL, edQTL, and sQTL in the AA cohort. Hypergeometric p values are shown to evaluate the overlap between each pair. **b,** Overlaps among the genes targeted by the 3 types of QTL in the AA cohort. Hypergeometric p values are shown to evaluate the overlap between each pair. **c,** Pearson's chi-square p values (Methods) between pairs of molecular targets (EL: editing level, PSI or RPKM) for loci shared by each pair of QTL in EU. Only loci with at least one significant pairwise correlation are shown (i.e., p value < 0.05).



Supplementary Figure 3-5. BLASTn alignment of regions harboring the seedQTL (rs146498205) and its target editing site.

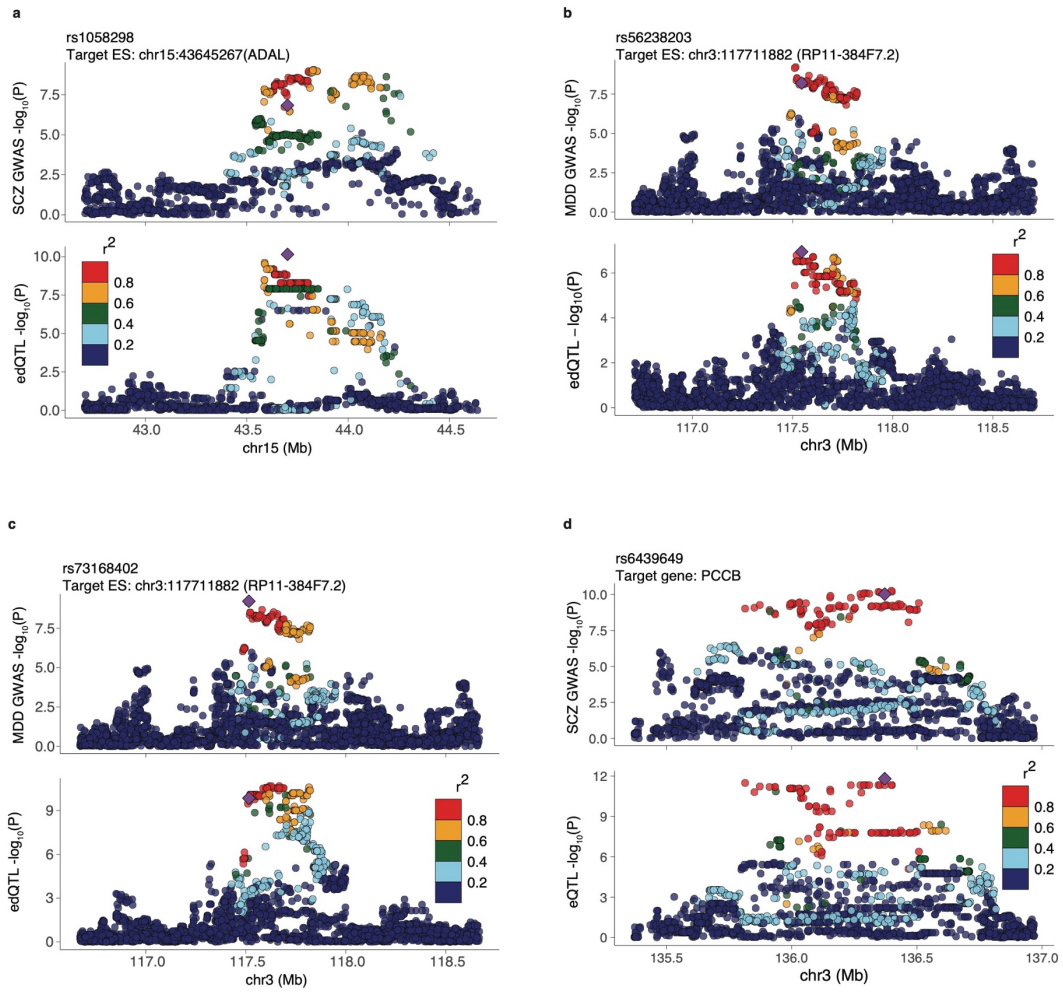
BLASTn alignment was conducted on two inverted ALUs containing a seedQTL (rs146498205) and its target editing site, respectively. Alignment results are shown between the reference allele of the seedQTL and the edited base of its target (top) and between the alternative allele of the seedQTL and the unedited base of its target (bottom).



Supplementary Figure 3-6. CTSB 3' UTR predicted structures.

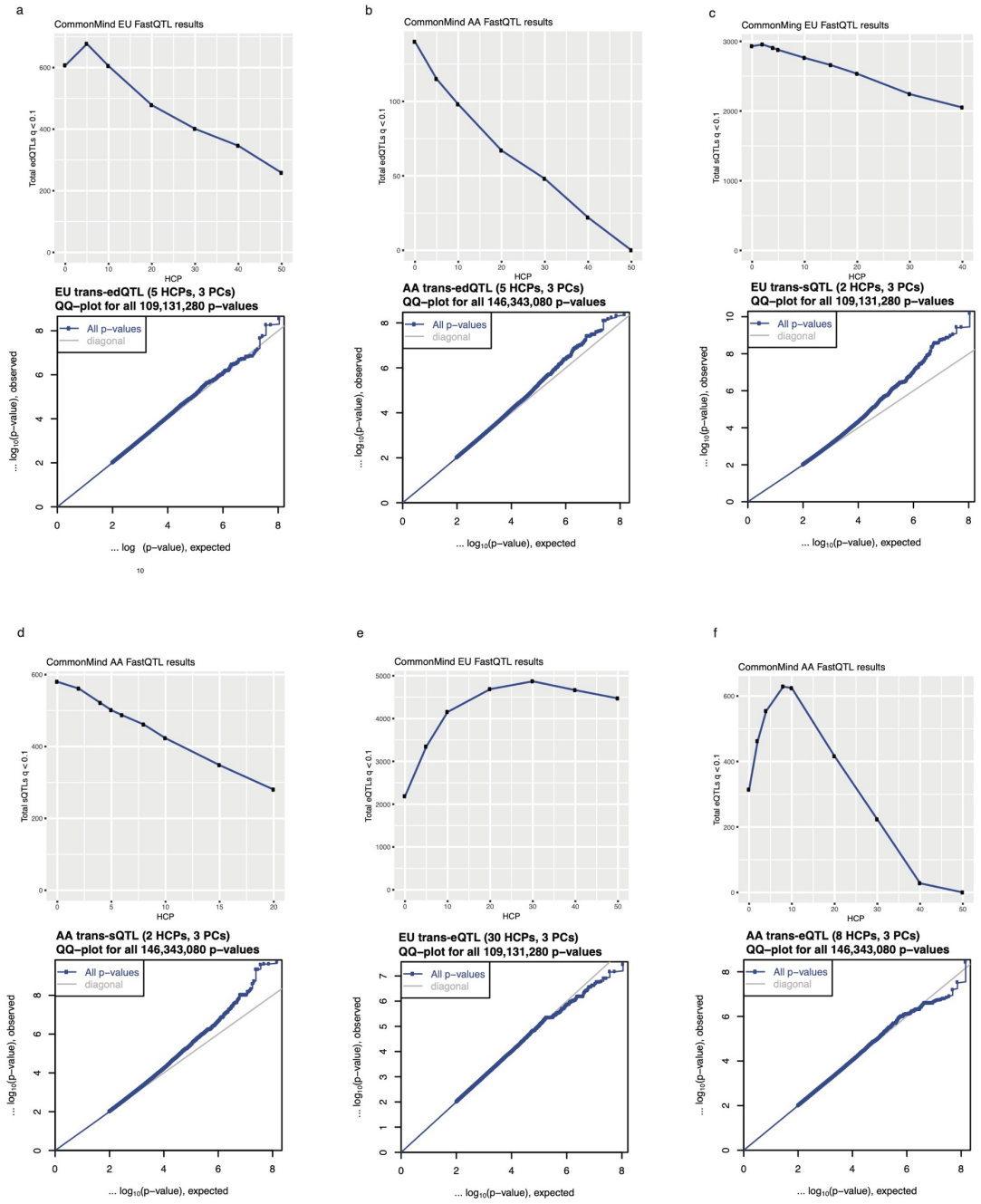
a, RNA structure of the CTSB 3'UTR predicted by RNAfold, given the reference allele of the edQTL, the edited allele at chr8:11700373, and the unedited allele at chr8:11700685. MFE: minimum free energy. **b,** Similar to (a), given the reference allele of the edQTL, the unedited allele at chr8:11700373, and the edited allele at chr8:11700685. **c,** Similar to (a), given the reference allele of the edQTL and both editing sites (ES) unedited. **d,** Similar to (a), given the reference allele of the edQTL and both ES edited. **e,** Similar to (a), given the alternative allele of the edQTL, the edited allele at

chr8:11700373, and the unedited allele at chr8:11700685. **f,** Similar to (a), given the alternative allele of the edQTL, the unedited allele at chr8:11700373, and the edited allele at chr8:11700685. **g,** Similar to (a), given the alternative allele of the edQTL and both ES unedited. **h,** Similar to (a), given the alternative allele of the edQTL and both ES edited.



Supplementary Figure 3-7. Additional QTL-GWAS colocalizations.

a, Example colocalization between SCZ GWAS (top) and an edQTL (bottom) targeting an editing site in the adenosine deaminase-like gene, ADAL. **b-c,** Example colocalizations between MDD GWAS (top) and two edQTL (bottom) targeting the same editing site in the lncRNA RP11-384F7.2. **d,** Example colocalization between SCZ GWAS (top) and an eQTL (bottom) targeting the mitochondrial gene PCCB.



Supplementary Figure 3-8. Selection of HCP and genotype PCs.

a, Total edQTL in EU detected by FastQTL for different hidden covariates (HCPs) included in the model (top). Population structure was evaluated by generating QQ plots of trans-associations using MatrixEQTL (bottom). The nonzero HCP value that yielded the most QTL from FastQTL and 3 genotype PCs were used in the final FastQTL model (Methods). **b,** Similar to (a), but for edQTL in AA. **c,** Similar to (a), but for sQTL in EU. **d,** Similar to (a), but for sQTL in AA. **e,** Similar to (a), but for eQTL in EU. **f,** Similar to (a), but for eQTL in AA.

3.11 TABLES

Ethnicity	QTL type	HCPs (n)	Genotype PCs (m)	Distance	Total QTL (FDR p <0.1)	Total QTL targets	Total QTL-target associations
EU	edQTL	5	3	100 KB	626	677 ES	677
AA	edQTL	5	3	100 KB	113	115 ES	115
EU	sQTL	2	3	1 MB	2,859	3,625 introns	3,626
AA	sQTL	2	3	1 MB	558	732 introns	732
EU	eQTL	30	3	1 MB	4,870	5,006 genes	5,017
AA	eQTL	8	3	1 MB	629	647 genes	647

Table 3-1. Overview of total QTL detected in EU and AA, respectively.

3.12 SUPPLEMENTARY TABLES

Supplementary Table 3-1. sQTL overlapping eCLIP-seq peaks of DDX3X and the associated target introns.

(Attached)

Disease	*Ethnicity*	*N case*	*N control*	*N*	*Source*
SCZ	European	52,017	75,889	127,906	Trubetskoy et al. (2022) Nature
SCZ	AA	5,998	3,826	9,824	Trubetskoy et al. (2022) Nature
MDD	European	170,756	329,443	500,199	Howard et al. (2019) Nat. Neuro.
BPD	European	24,972	449,593	474,217	Mullins et al. (2021) Nat. Gen.
ADHD	European	19,099	34,194	53,293	Demontis et al. (2019) Nat. Gen.
ASD	European	18,383	27,969	46,352	Grove et al. (2019) Nat. Gen.
PTSD	European	12,080	33,446	45,526	Nievergelt et al (2019) Nat. Comm.
PTSD	AA	4,363	10,976	15,339	Nievergelt et al (2019) Nat. Comm.
Alcohol Dependence	European	10,206	28,480	38,686	Walters et al (2018) Nat. Neuro.
Alcohol Dependence	AA	2,991	2,808	5,799	Walters et al (2018) Nat. Neuro.

Supplementary Table 3-2. Sources for GWAS summary statistics.

Supplementary Table 3-3 All significant QTL-GWAS colocalizations (PP H4 > 50).

(Attached)

CHAPTER 4 – Genome-wide alteration of dsRNA expression and RNA editing in Alzheimer's disease

4.1 ABSTRACT

Dysregulation of endogenous dsRNA expression and editing can trigger aberrant inflammatory responses within the cell. However, the profile, function and regulation of editing enriched dsRNAs in humans are unknown and have not been studied in most diseases. Here, we used known RNA editing sites and RNA structure prediction methodology to develop a pipeline for the detection of long dsRNAs. We identified 62,008 high-confidence editing enriched dsRNAs genome-wide, most of which overlap with introns and ALU repeat elements. The pipeline was applied to the dorsolateral prefrontal cortex (DLPFC) of individuals with Alzheimer's disease (AD) and controls from the ROSMAP consortium. We observed that the majority of differentially expressed dsRNAs were upregulated in AD and associated with increased interferon (IFN) stimulated gene expression. Furthermore, we observed reduced RNA editing at loci distinct from differential dsRNAs that were also associated with increased IFN response and AD severity. The results suggest that the global upregulation of dsRNAs and reduced RNA editing in nonoverlapping regions may collectively influence inflammatory response in AD.

4.2 INTRODUCTION

Long double-stranded RNAs (dsRNAs) are typically associated with viral propagation in mammalian cells and can trigger cellular response against infection[205]. Cytoplasmic dsRNA binding proteins (dsRBPs), such as DEXD/H-box helicase 58

(DDX58, also known as RIG-I) and melanoma differentiation-associated protein 5 (MDA5), detect and label such dsRNAs to be exogenous intruders within the cell, thus activating expression of type 1 interferons (IFN1) and other pro-inflammatory cytokines[206]. While dsRNAs originate in viruses, these structures are also endogenously produced in invertebrates and vertabrates[207–209]. Long endogenous dsRNAs are edited by adenosine deaminases (ADARs), a family of dsRBPs, that mark them as "self" RNAs[210–212]. RNA editing by ADARs converts adenosines to inosines (which are recognized as guanosines by downstream mechanisms), and enables dsRNAs to escape from dsRBP sensing and IFN activation[210–212], most likely by altering the dsRNA structure. In fact, the presence of A-to-I editing has been used as evidence for dsRNA structure formation, thus providing a means to detect dsRNAs genome-wide (referred to as the dsRNAome)[213]. Although detection of editing enriched regions (EERs) and prediction of their associated long dsRNAs have been carried out in previous literature[207,208,213], the total repertoire of long endogenous edited dsRNAs within the human transcriptome is largely unestablished.

Despite the fact that the function of the dsRNAome is poorly understood, dysregulation of dsRNAs, dsRBPs, and ADAR is known to cause aberrant recognition of endogenous transcripts, inducing IFN response[213]. Mutations in MDA5 may enhance its dsRNA binding capacity and can enable recognition of edited dsRNAs that normally escape from MDA5 sensing. Such MDA5 mutations have been associated with immune hypersensitivity in the Aicardi–Goutières syndrome (AGS)[214]. In addition to intramolecular dsRNAs, intermolecular dsRNAs can form between sense-antisense transcripts, such as between the transcript for β secretase-1 (BACE1) and its natural antisense transcript (BACE1-AS) in the context of Alzheimer's disease (AD)[215]. Indeed, dysregulated dsRNA

expression, such as dsRNA accumulation, and abnormal editing have been associated with autoimmune and neurodegenerative disorders like AGS, ALS, and AD[75,212,215–218]. However, the global characterization of these dsRNAs and their related RNA editing profiles has not been conducted in most immune and neurological diseases.

AD is a neurodegenerative brain disorder characterized by the toxic accumulation of amyloid-beta peptide (Aβ) and tau proteins, which lead to memory loss, cognitive decline, and death associated with aging[219,220]. Furthermore, dysregulated inflammatory processes, such as proinflammatory cytokines and IFNγ stimulation, have been recognized as distinct contributors of AD progression[221]. Thus far, only a few studies have reported RNA editing dysregulation and specific differentially expressed dsRNAs in AD[75,124,215,222,223]. These works motivate global dsRNA and RNA editing detection in AD to further understand the differential dsRNAome in the disease.

Here, we developed a pipeline to detect endogenously expressed, editing enriched, long dsRNAs. We applied the method to RNA-seq data of the dorsolateral prefrontal cortex (DLPFC) of AD and control samples (184 and 163 samples, respectively) from the Religious Orders Study (ROS) and Rush Memory and Aging Project (MAP), referred to as ROSMAP[69]. Given the known role of RNA editing in modulating dsRNA sensing, we also investigated differential RNA editing in AD using our in-house protocols. Both differential RNA editing sites and differentially expressed dsRNAs occurred in repetitive elements known as ALUs, and the majority fell into introns and intergenic regions. We observed that editing was significantly lower in AD patients and the majority of differential dsRNAs were upregulated in AD relative to controls. Despite dysregulated RNA editing and dsRNA expression occurring in distinct regions, both are significantly

associated with interferon stimulated genes (ISGs). To our knowledge, our study presents the first comprehensive bioinformatic dsRNA detection and characterization in AD and its relationship to RNA editing. Our results suggest that dysregulation of RNA editing and dsRNA expression in nonoverlapping regions may synchronously contribute to increased immune response in AD.

4.3 RESULTS

4.3.1 Overview of dsRNA detection

We expanded upon a previously published method to detect editing enriched regions (EERs)[208] and developed a pipeline to identify and characterize dsRNAs genome-wide based on RNA editing sites from REDIportal[88] (Methods). An overview of the pipeline is presented in Fig. 4-1. Briefly, we merged editing enriched 50 bp sliding windows that were $\leq$ 1kb apart. We retained those EERs between 300 bp and 4 kb (Fig. 4-1) and found that most EERs identified were < 1 kb in length (Fig. 4-2a). We obtained high-confidence dsRNAs through structure annotations by utilizing both the RNAfold software[168] and additional methods developed in our lab (Fig. 4-1, Methods). The average minimum free energy (AMFE) value of each dsRNA was used to determine structure stability (Methods).

As expected, the majority of EERs were located in ALU regions (98% with AMFE < -0.35, Fig. 4-2b). Furthermore, most EERs were in introns or intergenic regions (41% and 36% with AMFE < -0.35, respectively). This was followed by 12% of EERs spanning both introns and exons (AMFE < -0.35, Fig. 4-2c). We empirically selected an AMFE cutoff of < -0.35 for dsRNAs to be considered “high-confidence” predictions. This cutoff was determined through examination of the overall AMFE distribution (Fig. 4-2d) and example

dsRNA structures from RNAfold[168] (Fig. 4-2e, Supplementary Fig. 4-1a, b). After applying the AMFE cutoff, a total of 62,008 dsRNAs were retained for downstream analyses. Fig. 4-2e demonstrates a predicted dsRNA with the lowest AMFE value and most stable structure. Example dsRNAs with less stability (higher AMFE values) are presented to exemplify the range of structures predicted through the dsRNA pipeline (Supplementary Fig. 4-1a, b).

4.3.2 Differential dsRNA detection and IFN response in AD

We obtained RNA-seq data from the DLPFC region of control and AD ROSMAP samples[69,224,225]. A total of 633 samples were mapped using hisat2[126], and unmapped reads were realigned with a pipeline developed to resolve mapping of hyper-edited RNA-seq reads[41,87] (Methods). To detect dsRNAs that were differentially expressed in AD, we used CERAD scores, which quantify neuritic plaques within the brain[70], to choose subsets of samples for differential comparison. Samples with CERAD scores 1 and 4 were labeled as AD (n = 184) and controls (n = 163), respectively. We utilized the HTSeq[134] software and dsRNA coordinates to obtain raw read counts for all regions. The Limma-Voom[226] package was applied next, which identified a total of 1,566 differentially expressed dsRNAs between AD and controls (FDR adjusted $P < 0.05$, Methods). Interestingly, we found that 1,097 (70%) out of all differential dsRNAs had upregulated expression in AD (Fig. 4-3a).

As dsRNA upregulation is known to elicit innate immune response[205,227], we asked whether differential dsRNAs detected in AD may induce ISG expression. We correlated the expression of 1,097 upregulated dsRNAs with that of known IFN-induced genes[228]

using a linear regression model, including age and sex as covariates (Methods). We observed a clear pattern of positive correlation between almost all upregulated dsRNAs and the majority of ISGs (Fig. 4-3b). This observation supports that increased dsRNA expression in AD may lead to increased targeting of these transcripts by immune response dsRNA sensors, such as MDA5[205,210,227]. Therefore, upregulated dsRNA expression may be one underlying cause for increased immune response observed within AD[229,230].

4.3.3 Differential RNA editing in AD

As RNA editing is known to occur in endogenous dsRNAs and modulate their immunogenicity, we were motivated to investigate differential editing in AD. We analyzed RNA editing using two methodologies: 1) detecting *de novo* differential RNA editing in AD using previously developed protocols[41,149], and 2) associating ALU editing index (AEI) with AD severity. For the purpose of differential RNA editing detection between AD and controls, we grouped the samples by CERAD scores, as conducted in previous literature[231,232]. Those with little to no AD pathology (CERAD scores 3 and 4) were considered as control samples, while those with high AD pathology (CERAD scores 1 and 2) were labeled as AD. We conducted quality control methods described previously[41,149] to ensure that no metadata covariates differed between conditions (Supplementary Fig. 4-2). After the data processing and quality control procedures, we retained 415 samples (202 AD and 213 controls) for downstream differential RNA editing analysis.

We applied our RNA editing detection pipeline[84,127,128] and identified 3,131,119 RNA-DNA differences in all samples (Methods). A total of 556,688 loci occurring in at least 10% of samples were referred to as “common” RNA editing sites and used for downstream analyses (Methods). Over 95% of editing sites detected per sample were A-to-G sites (Supplementary Fig. 4-3a), demonstrating the high accuracy of our pipeline. Furthermore, the vast majority of common sites were in ALU regions (Supplementary Fig. 4-3b). We identified 4,438 RNA editing sites with either significant difference in editing ratios or editing prevalence between AD and controls (Methods, Supplementary Fig. 4-4a). Differential sites were primarily located in intergenic and intronic regions, followed by the 3’UTRs (Supplementary Fig. 4-3c).

We calculated differential RNA editing average (DREA) for all samples (Methods), and found that AD samples have significantly reduced editing compared to controls (Supplementary Fig. 4-4b). Based on the editing levels for all differential sites, samples were generally clustered by disease condition (Supplementary Fig. 4-4c). We identified 8 differential editing sites causing nonsynonymous change to their proteins, hereby referred to as “recoding” sites. The majority of recoding sites (5 out of 8) showed lower average editing in AD (Supplementary Fig. 4-4d), including an A-to-G recoding site in GRIA4, a glutamate receptor associated with nicotine dependence and major depression[233]. To examine functional enrichment of differential loci, we conducted gene ontology (GO) analysis of differentially edited genes. This yielded interesting pathways related to brain function and immune response, such as synaptic plasticity, antigen processing and presentation via MHC class II, and neuromuscular process (Supplementary Fig. 4-4e).

Finally, we asked whether differential RNA editing sites overlapped differentially expressed dsRNAs. We found that only 2% of the differential dsRNAs harbored differentially edited loci (1% of the differential editing sites), which was insignificant when compared to random common RNA editing sites in ROSMAP (Chi-squared P = 0.90). Thus, the differential RNA editing sites and differential dsRNAs in AD appeared to reside in distinct regions. Next, we asked whether differential RNA editing sites in our analysis (showing global hypo-editing) also correlated with ISG stimulation, as seen for upregulated dsRNAs. Indeed, we observed that DREA from these distinct RNA editing loci frequently showed significantly negative correlation with ISGs (Supplementary Fig. 4-4f).

4.3.4 ALU editing index associated with AD

Next, we aimed to examine the global hypo-editing trend relative to AD severity and the distinction between differentially edited loci and differentially expressed dsRNAs. To this end, we analyzed RNA editing with the metric ALU editing index (AEI) using the RNAEditingIndexer tool[234]. Specifically, AEI summarizes editing levels within all ALU regions. We associated AEI with AD severity through ordinal regression (Methods). AD severity was determined via alternative metrics, including Braak stage, cognitive index, and CERAD score. While Braak stage refers to cognitive decline measured via neurofibrillary tangles in the brain[71], CERAD score is a semiquantitative measure of neuritic plaques[70]. The cognitive index differs from these two metrics as it combines the results of 19 different neuropsychological tests into a single “Global Cognitive Score”[69].

Ordinal regression of each ALU's AEI value with AD severity was conducted for all 633 samples, including age and sex in the model (Methods). We observed that the majority of ALU regions that were significantly associated with at least one AD metric (FDR adjusted $P < 0.05$), had negative correlation between editing and AD severity (Fig. 4-4a). This observation is consistent with our findings of hypoediting of individual editing sites in AD in the previous section. We then conducted GO enrichment analyses of two groups of genes: those containing ALUs negatively correlated with AD severity and those with positive correlations. The first group yielded the most interesting functional pathways, such as mitochondrial inner membrane, post-synaptic density, and skeletal muscle tissue development (Fig. 4-4b). These terms were similar to those associated with differential editing in other brain disorders, such as SCZ[41,149].

Finally, we examined whether ALUs associated with AD severity fell into differentially expressed dsRNAs. Consistent with our observation with the individual RNA editing sites, we found that only two differential dsRNAs contained differentially edited ALUs (< 1% of ALUs and dsRNAs). Therefore, this analysis confirms that hypo-editing in AD occurs in dsRNA forming regions that are distinct from those dsRNAs differentially expressed in the disease. We infer that dysregulated RNA editing, which is known to induce aberrant immune response, and upregulated dsRNA expression may be two independent molecular mechanisms correlated with higher IFN in AD.

4.4 DISCUSSION

We developed a pipeline to identify genome-wide dsRNAs, which can be applied to a wide range of diseases. Through dsRNA detection in AD, we identified 1,566 differentially expressed dsRNAs, the majority of which were upregulated in the disease and positively associated with ISG expression. As RNA editing occurs in dsRNA regions, we further conducted differential RNA editing analysis and overlapped these loci with aberrantly expressed dsRNAs. We showed a global reduction in RNA editing in AD DLPFC and found that differential editing occurred in regions distinct from upregulated and downregulated dsRNAs.

In our comprehensive identification of dsRNA-forming structures, we focused on long transcripts that can be targeted by dsRBP sensors for IFN response, such as MDA5. As expected, most dsRNAs fell into ALU elements, as ALU:ALU hybrids were reported to serve as the primary endogenous ligands for MDA5 activation[214]. Consistent with previous literature, we observed that the majority of dsRNAs were located in intronic and intergenic regions[213,235]. Intronic dsRNAs may have been observed in the RNA-seq data due to intron retention associated with the disease[236]. Furthermore, intron retention in AD often occurs within genes involved in innate immune response[236]. It is possible that retained introns harbor immunogenic dsRNAs detected by dsRBPs. In general, accumulation of intronic and intergenic inverted retroelements (IIIRs) can activate endogenous dsRNA responses including OAS3/RNase L and PKR[235]. These immune sensors, which are distinct from the MDA5 and RIG-I IFN activation pathways, act as monitors for nuclear RNA decay through detection of the cytosolic escape of IIIRs[235].

Therefore, increased intron retention may be associated with multiple possible explanations for increased dsRNA expression observed in AD.

Previous analyses have detected differential RNA editing in AD brains[75,222,223]. While it is possible that lower editing in AD may lead to upregulation in dsRNAs, especially those dsRNAs residing in ISGs, we found that differential dsRNAs did not significantly overlap ISGs compared to random dsRNAs (Chi-squared P = 0.29). Therefore, we propose that reduced RNA editing and increased expression of dsRNAs are two independent (or indirectly linked) processes that can both lead to IFN response observed in AD (Fig. 4-5). The interplay between reduced RNA editing and aberrant dsRNA levels is complex. For example, RNA editing and dsRNA expression dysregulation may occur in distinct cell types, the effects of which cannot be disentangled using bulk RNA-seq. Alternatively, increased dsRNA expression at certain loci may strain the editing machinery, such as ADAR, leading to hypo-editing in other transcripts. Furthermore, deciphering the complex pathways by which these dysregulated molecules activate innate immune response by dsRNA sensors, such as MDA5, RIG-I, PKR, and TRL3, is challenging[237]. Therefore, experimental examination of the interplay between reduced editing and upregulation of dsRNAs in relation to ISG expression is necessary to provide further mechanistic insight in the future.

In summary, we present the first study to show globally reduced RNA editing and dsRNA accumulation within AD, and predict its collective impact on IFN response in the disease. Our work motivates future studies to examine the contribution of these two mechanisms to the innate immunity changes in AD.

4.5 METHODS

4.5.1 dsRNA detection pipeline

Building upon a method to identify EERs[208], we developed a pipeline to detect long editing enriched dsRNAs genome-wide. Our dsRNA detection pipeline included six steps outlined in Fig. 4-1. First, we obtained 15,683,855 human RNA editing sites from the REDIportal database[88] for the purpose of identifying loci with clusters of editing. To this end, we determined EERs by implementing a 50 bp sliding window across the genome, and aggregating those with ≥ 3 RNA editing sites. EERs within 1 kb of each other were merged together. This length was chosen empirically by determining a distance that would allow for merging of EERs within the same transcript, but not those in separate transcripts. Additionally, larger merge distances yielded longer dsRNAs that would sacrifice downstream structure prediction accuracy by RNAfold[238,239].

To exclude EERs that were too short for MDA5 recognition[240] or too long for accurate structure prediction[238], we next applied minimum and maximum length thresholds to merged EERs (Fig. 4-1). We chose to retain regions between 300 bp - 4 kb in length, which were cutoffs determined both empirically and from previous literature[240]. Edited dsRNAs that can be targeted by immune response proteins of interest, such as MDA5, must have long stem regions. These dsRBPs prefer dsRNAs > 2 kb, but can bind to dsRNAs as short as 300 bp in length[240]. Upon investigation of EER length distribution, we found that most EERs were < 1 kb (Fig. 4-2a). We chose a 4KB maximum length, as we were able to retain most EERs (>95%) with little compromise to RNAfold prediction accuracy[168].

We determined high confidence dsRNAs by applying the RNAfold software from the Vienna Package (v2.0)[168] to all EERs. We obtained minimum free energy (MFE) values and dot-bracket notations for the predicted RNA secondary structures with highest-confidence. Dot-bracket notations were interpreted for each nucleotide of a dsRNA, labeling stem, loop, single stranded, and bulge regions. Only EERs with a stem length of at least 200 bp containing < 20% mismatches were retained as dsRNAs. DsRNAs were overlapped with ALUs in RepeatMasker[132] and gene regions were annotated using ENSEMBL v99[130]. We calculated AMFE values by dividing RNAfold MFE by dsRNA length. An AMFE cutoff was determined by investigating AMFE distribution (Fig. 4-2d) and visualizing RNA structures with various AMFEs (Fig. 4-2e, Supplementary Fig. 4-1a, b). All dsRNAs with AMFE < -0.35 were retained to reach the final list of 62,008 high-confidence dsRNAs.

4.5.2 RNA-seq data analysis

We obtained stranded poly-A selected RNA-seq data from the DLPFC brain region of AD and controls from ROSMAP[224,225]. A total of 634 samples were obtained, which constitute the largest batch of the ROSMAP data[225]. The RNA-seq data were extracted as bam files originally mapped via STAR[241]. However, to maintain consistency with our previous methodologies[41,149], we ran the bamtofastq function from BedTools (v 2.3.0)[242] to convert all reads into unmapped fastq files. The fastq files were then mapped using hisat2 (v 2.0.4)[135], with default parameters other than not allowing discordant reads. Unmapped reads were re-aligned to regions enriched with editing sites that are usually

missed by standard mapping procedures[41,87]. Uniquely mapped reads were extracted and combined for downstream analyses.

4.5.3 Differential dsRNA detection in AD

We obtained read counts for all dsRNA regions using the HTSeq software (v 0.6.1)[134]. For differential analysis, we used only samples with extreme CERAD scores '1' and '4' (AD and Control, respectively) which may possess the most differences in expression. A total of 347 samples were retained (184 AD and 163 controls). DsRNAs with ≥ 2 reads in 20% of the 347 samples and those with $\geq 20\%$ difference in effect size (i.e. mean difference in CPM between AD and control samples) were considered for differential dsRNA expression analysis through the Limma-Voom software (v 3.46.0)[226]. Briefly, the model.matrix() function was used to create a linear model including sample condition, age, and sex. The voom() function was used to transform raw counts to logCPM and perform mean-variance modeling before lmFit() was applied to fit the linear model. Finally, the significance of differential dsRNA expression was tested using the eBayes() function. DsRNAs with FDR < 0.05 were defined as differentially expressed between AD and controls. Positive or negative log fold change values were utilized to determine whether a dsRNA's expression was upregulated or downregulated, respectively, in AD.

4.5.4 Differential dsRNA expression associated with interferon response

All significantly upregulated dsRNAs in AD were tested for association with known ISGs from previous literature[228]. DsRNAs and IFN genes were required to be expressed (FPKM > 0) in at least 20% of samples. Log dsRNA FPKM was correlated to log IFN gene

FPKM by fitting a linear model for each ISG with normalized dsRNA expression, sex, and age of death of the individual sample as covariates. The square-root of the correlation (R) and P value of the resulting model were used to summarize the association significance. To visualize correlation directionality between upregulated dsRNAs and IFN response genes, the resulting R was made positive or negative to be consistent with the positive or negative dsRNA expression coefficient in the model.

4.5.5 Differential RNA editing in AD

We detected RNA-DNA differences (RDDs) using our previously developed methodologies[84,127,128]. We retained loci that had at least 2 edited reads and 5 total reads in at least 10% of samples. These sites were referred to as “common” RNA editing sites, and were used for downstream analyses.

We labeled samples as AD and Control for differential editing detection using CERAD scores, which is commonly used to categorize AD in the literature[231,232]. Those with CERAD scores ‘1’ and ‘2’ were categorized as AD samples, while those with ‘3’ and ‘4’ were considered as controls. We conducted quality control (QC) procedures to ensure that AD and control samples did not have a significant difference in any meta data variables, such as age, gender, RIN, and other mapping metrics[41,149] (Supplementary Fig. 4-2). A total of 415 samples (202 AD and 213 controls) passed our QC analysis.

We used a method previously developed in our lab that uncovers sites with difference in editing ratio or editing prevalence between AD and controls[41,149]. This method allows a flexible read coverage requirement to adapt to the different total read coverage available per site[41,149]. Briefly, we identified the highest coverage (C) possible (between ≥20, 15, or 5 reads) for each site (e_i) with a minimum of 5 samples per condition.

Average editing level per condition (A_{iAD} and $A_{iControl}$) was calculated using samples with a minimum of *C* coverage. Samples that fulfilled lower read coverage thresholds (≥15, 10, or 5) were considered and included in A_{iAD} or $A_{iControl}$ if their inclusion did not alter the average editing level by > 0.03. Occasionally, an initial read coverage requirement *C* was not reached due low read coverage for too many samples. In this case, we calculated average editing level per condition for e_i using samples with ≥5 read coverage, as long as 20% of samples from each condition reached this coverage requirement. Significant editing difference in AD and control groups were determined via Wilcoxon rank-sum tests. Differential editing sites were those with Wilcoxon rank-sum $P < 0.05$ and an effect size > 5%.

Fisher's exact tests were conducted to compare the total numbers of AD and control samples with non-zero editing level vs. zero editing level (using the adaptive read coverage requirements described above)[41]. Those sites with $P < 0.05$ and effect size > 5% were considered to have differential prevalence in editing between conditions.

We calculated differential RNA editing average (DREA) per sample using the mean editing level across all differential editing sites with read coverage ≥ 5 reads. To determine the significance of overall DREA between AD and controls, Wilcoxon rank-sum test was conducted. A heatmap of differential RNA editing levels per sample was generated via the R package gplots[133], similarly to that generated in our previous work[149].

4.5.6 DREA association with ISG expression

We calculated the association between global differential editing levels and the expression of ISGs. Spearman correlation was conducted between sample DREA and

expression levels of IFN stimulated genes[228]. Spearman ρ coefficients and P values were used to evaluate correlation directionality and significance.

4.5.7 AEI associated with AD severity

We conducted ordinal regression analysis between ALU editing index (AEI) and pathological measures in AD, including CERAD score, cognitive index, and Braak stage. We ran the RNAEditingIndexer[234] software with default parameters to generate read pileups for all adenosines within each ALU region annotated by RepeatMasker[132]. Briefly, mismatches and matches (i.e. G and A in plus strand reads, respectively; C and T in minus strand reads, respectively) for each base within an ALU were obtained. For each ALU, the sum of all mismatches within its start and end coordinate was divided by the sum of total mismatches and matches and then multiplied by 100 to obtain AEI. ALU AEI values were scaled so that each ALU had mean = 0 and variance = 1 across samples. We conducted ordinal regression between each disease pathological measure and corresponding sample AEI values, including age of death and sex as model covariates. ALUs were required to be expressed in at least 25% of samples in each CERAD group to be included in the ordinal analysis. Ordinal regression was conducted using the clm() function in the ordinal R package[243]. P values and the AEI coefficient obtained from

ordinal regression were plotted using the ComplexHeatmap R software[244]. AEI coefficients with p-value < 1e-4 were considered significant.

4.5.8 GO enrichment

GO enrichment was conducted using methodologies previously developed[41,149], with the exception that genes matched for expression were determined via average expression calculated from all ROSMAP samples.

4.6 ACKNOWLEDGEMENTS

This work was supported in part by grants from the National Institutes of Health (R01MH123177, R01AG075206 to X.X.). M.C. was supported by the NIH T32LM012424. E.H was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2034835.

4.7 AUTHOR CONTRIBUTIONS

M.C., E.H. and X.X. designed the study. M.C. developed the pipeline for dsRNA detection and conducted differential RNA editing analysis. E.H. and M.C. identified differential dsRNAs. E.H conducted analyses with AEI and AD severity. X.X. provided supervisory inputs and was the principal investigator.

4.8 COMPETING INTERESTS

The authors declare no competing interests.

4.9 FIGURES

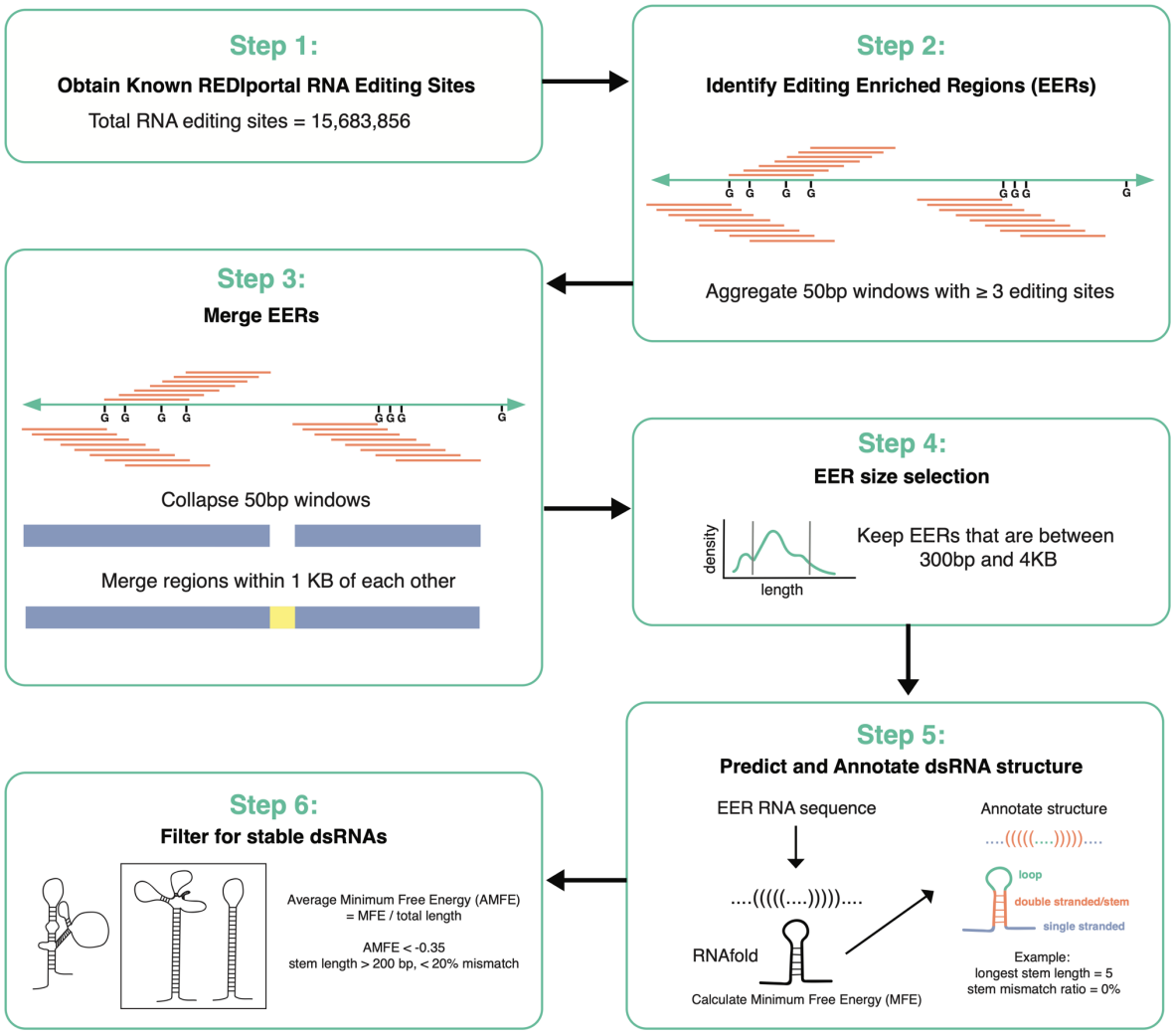


Figure 4-1. dsRNA detection pipeline.

RNA Editing sites from REDIportal were extracted (step 1) and 50bp sliding windows throughout the genome with ≥3 editing sites were aggregated (step 2). All windows within 1KB of each other were merged (step 3), and those between 300 bp and 4 kb where retained (step 4). We applied the RNAfold software to get dot and bracket RNA structure notation for each EER along with average minimum free energy (AMFE) (step 5, Methods). Dot and bracket notation was translated into RNA structure annotations for loop, stem, bulge, or single-stranded regions (step 5). Stable "high-confidence" dsRNA structures were defined as those with stem length > 200bp, stem nucleotide mismatch ratio < 20%, and AMFE < -0.35 (step 6).

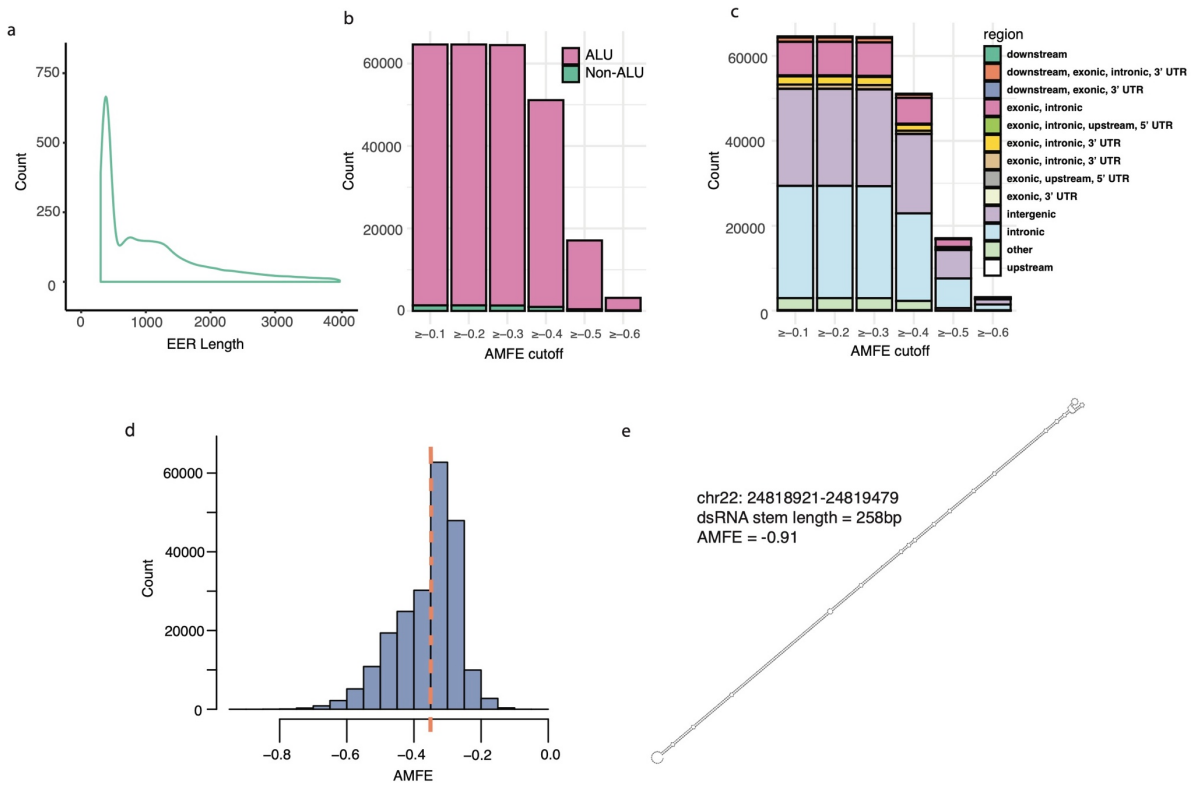


Figure 4-2. Overview of predicted dsRNAs.

a, Distribution of EER length after minimum and maximum thresholds were applied (300 bp and 4 kb, respectively). **b,** Total count of EERs overlapping (or not overlapping) ALU regions, separated by applied AMFE threshold. **c,** Total count of EERs overlapping genomic regions, separated by applied AMFE threshold. Upstream and downstream labels refer to regions 2 kb before the transcription start site and after the transcription end site of an annotated gene, respectively. **d,** Distribution of AMFE values for all EERs (between 300 bp and 4 kb in length) tested by RNAfold. Lower AMFE values represent more stable dsRNA structures. **e,** Example dsRNA with one of the most stable structures according to AMFE and dsRNA stem length.

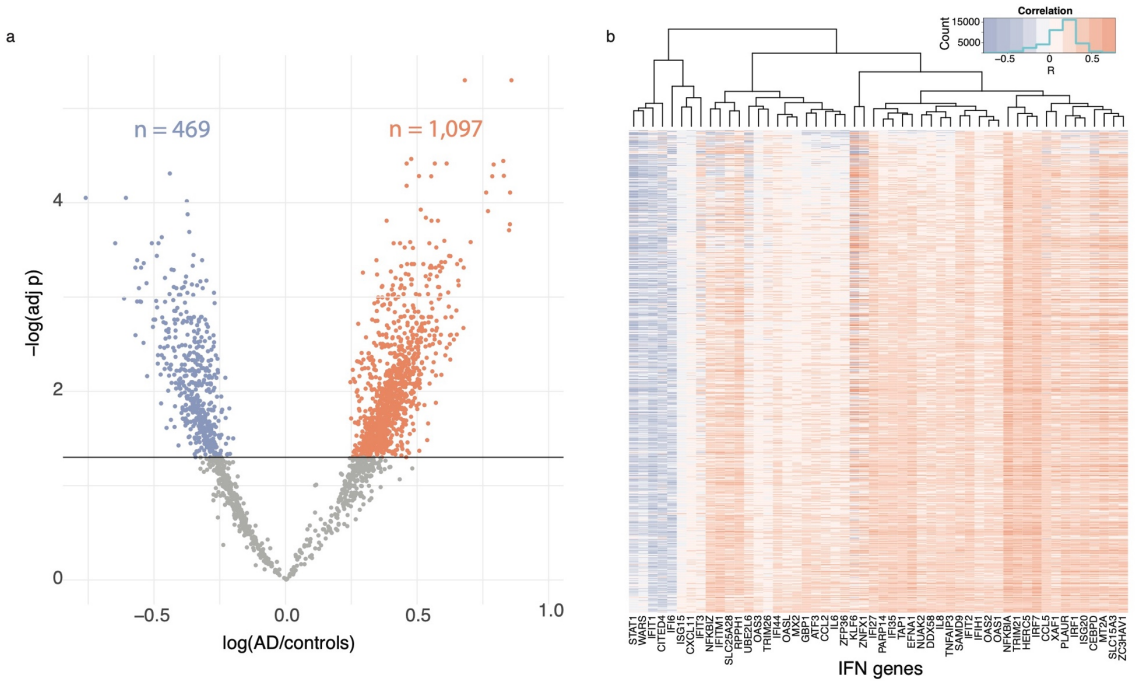


Figure 4-3. Differential dsRNA expression associated with IFN response.

a, The significance and fold change (FC) of dsRNA expression in AD relative to controls, as calculated by LimmaVoom. Orange and purple points represent dsRNAs with significantly upregulated and downregulated expression in AD, respectively (adjusted $p < 0.05$). **b,** Correlation between 1,097 significantly upregulated dsRNAs in AD and expression of known ISGs. Correlation coefficients (R) were calculated via a linear model associating dsRNA expression to each ISG, including age and sex as covariates (Methods).

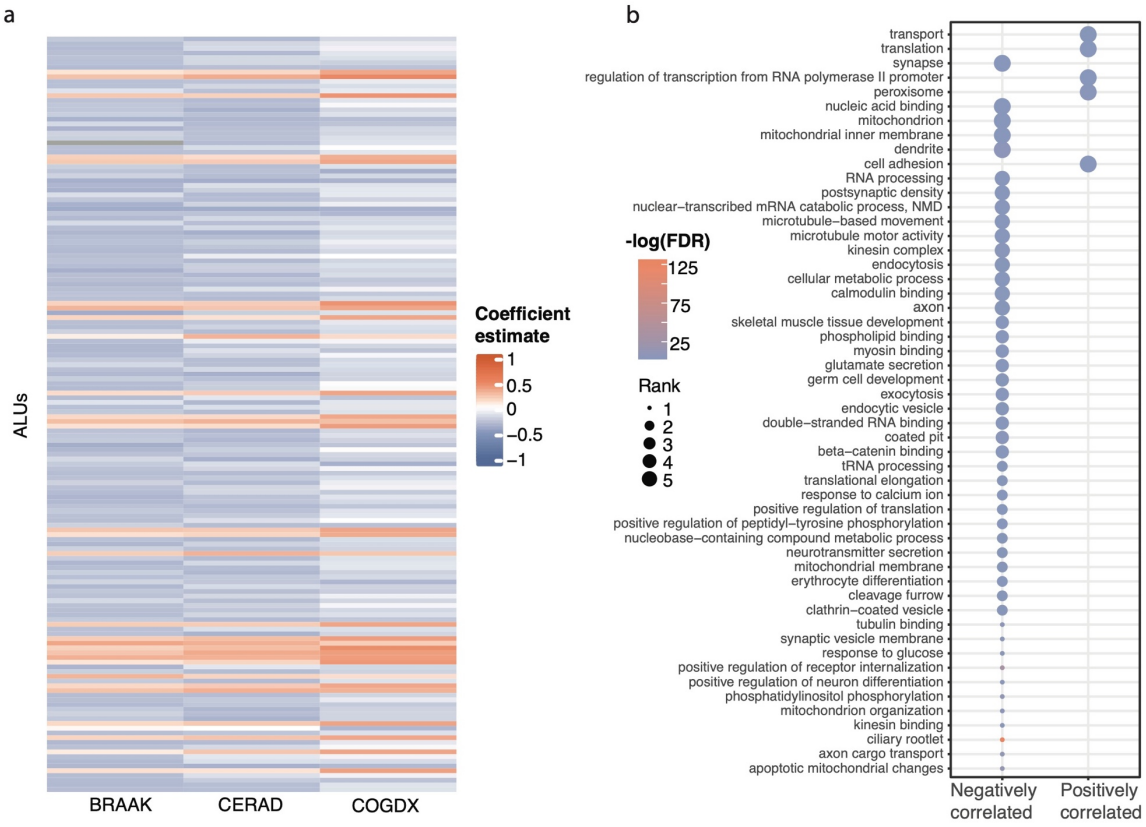


Figure 4-4. Differential AEI and functional enrichment.

a, ALU AEI associated with all three measures for AD severity. Coefficient estimates were obtained from ordinal regression of each ALU's AEI over the AD measure of interest, including age and sex as covariates. Only ALUs with significant correlation (P < 1e-4) in at least one AD measure are plotted. **b,** GO enrichment of genes with ALUs that were negatively or positively correlated (correlation estimate < 0 or > 0, respectively) across all AD measures and reached significance (P < 1e-4) for at least one AD measure.

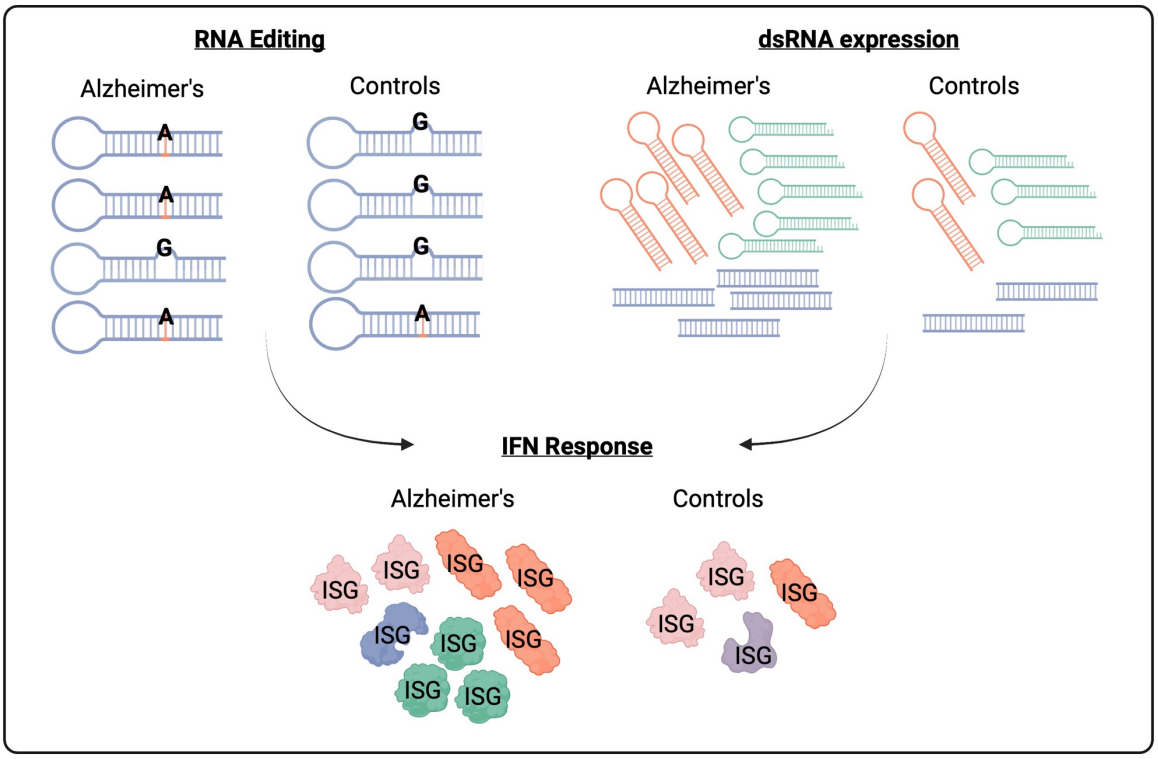
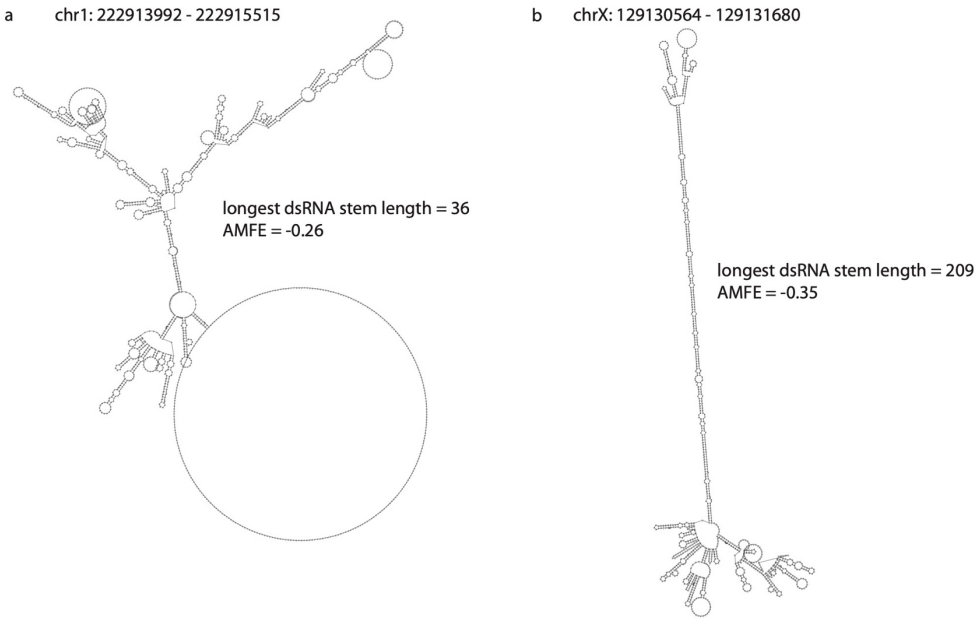


Figure 4-5. Overview of dysregulated editing, dsRNA expression, and IFN response in AD.

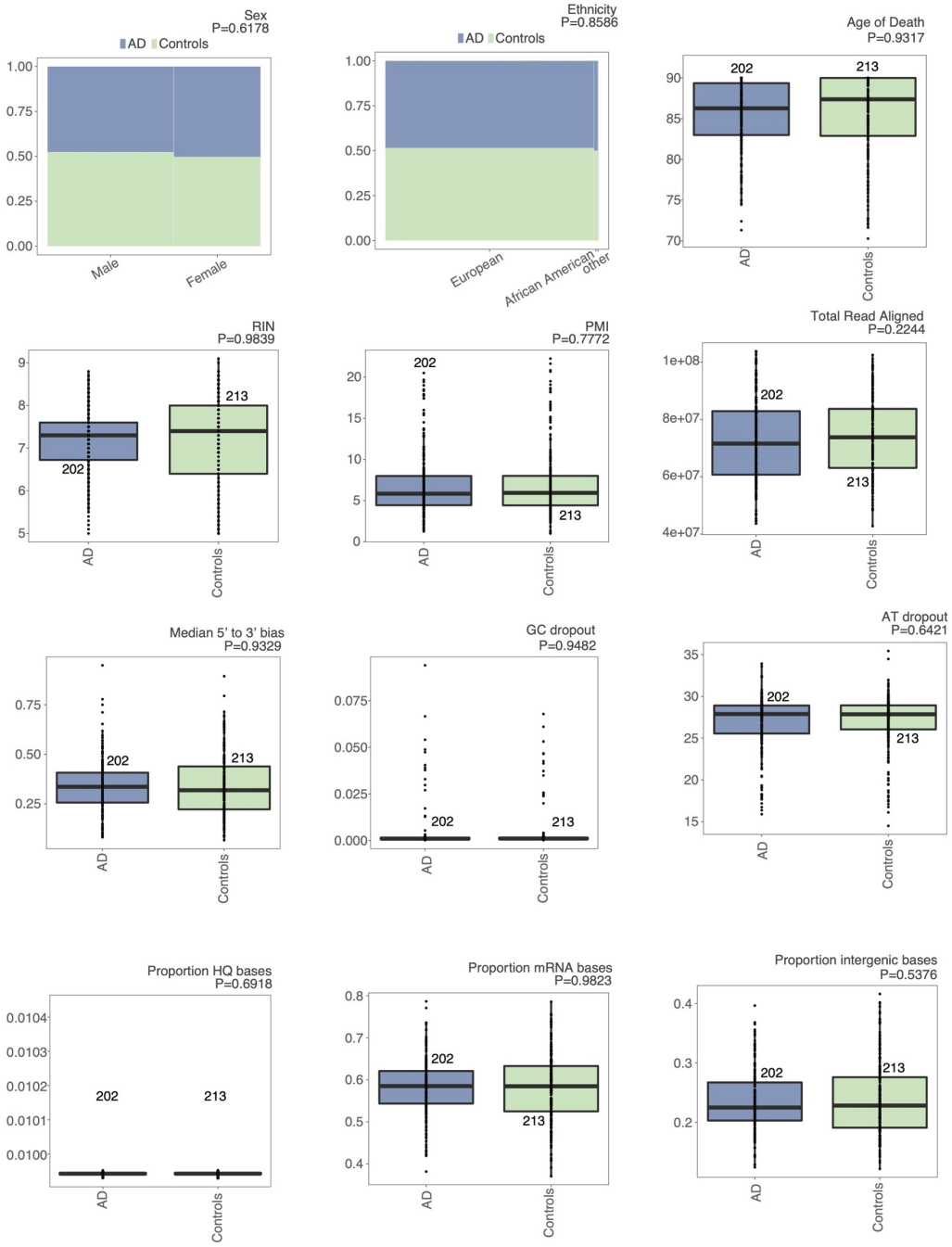
Reduced RNA editing and increased dsRNA expression in AD occur in separate transcripts. Nevertheless, both mechanisms significantly associate with increased IFN response in the disease.

4.10 SUPPLEMENTARY FIGURES



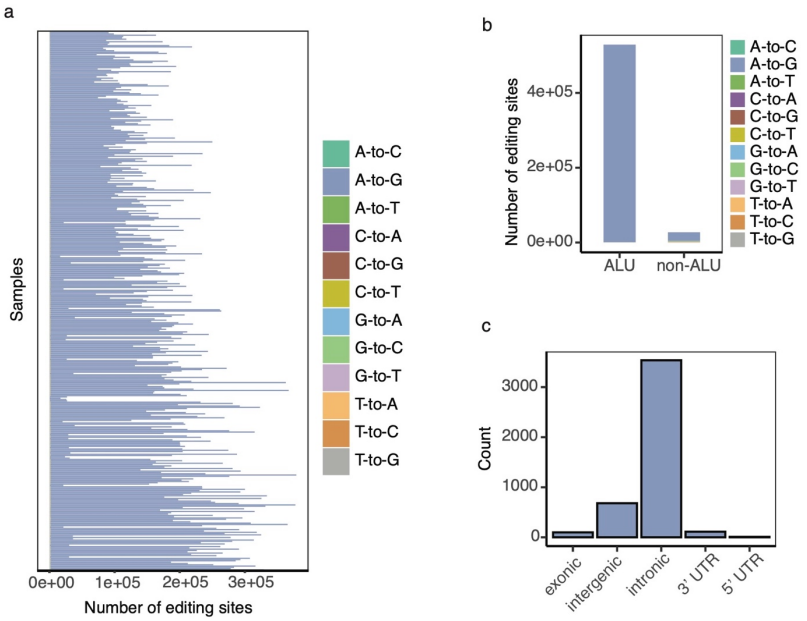
Supplementary Figure 4-1. Example EER structures predicted by RNAfold.

a, Example of a weak EER structure predicted by RNAfold. This region was considered unstable as the AMFE was > -0.35 and the stem length was < 200 bp. **b,** Example of a predicted structure from RNAfold that passed the minimum requirements to be considered a “high-confidence” long dsRNA region (AMFE < -0.35, stem length > 200bp).



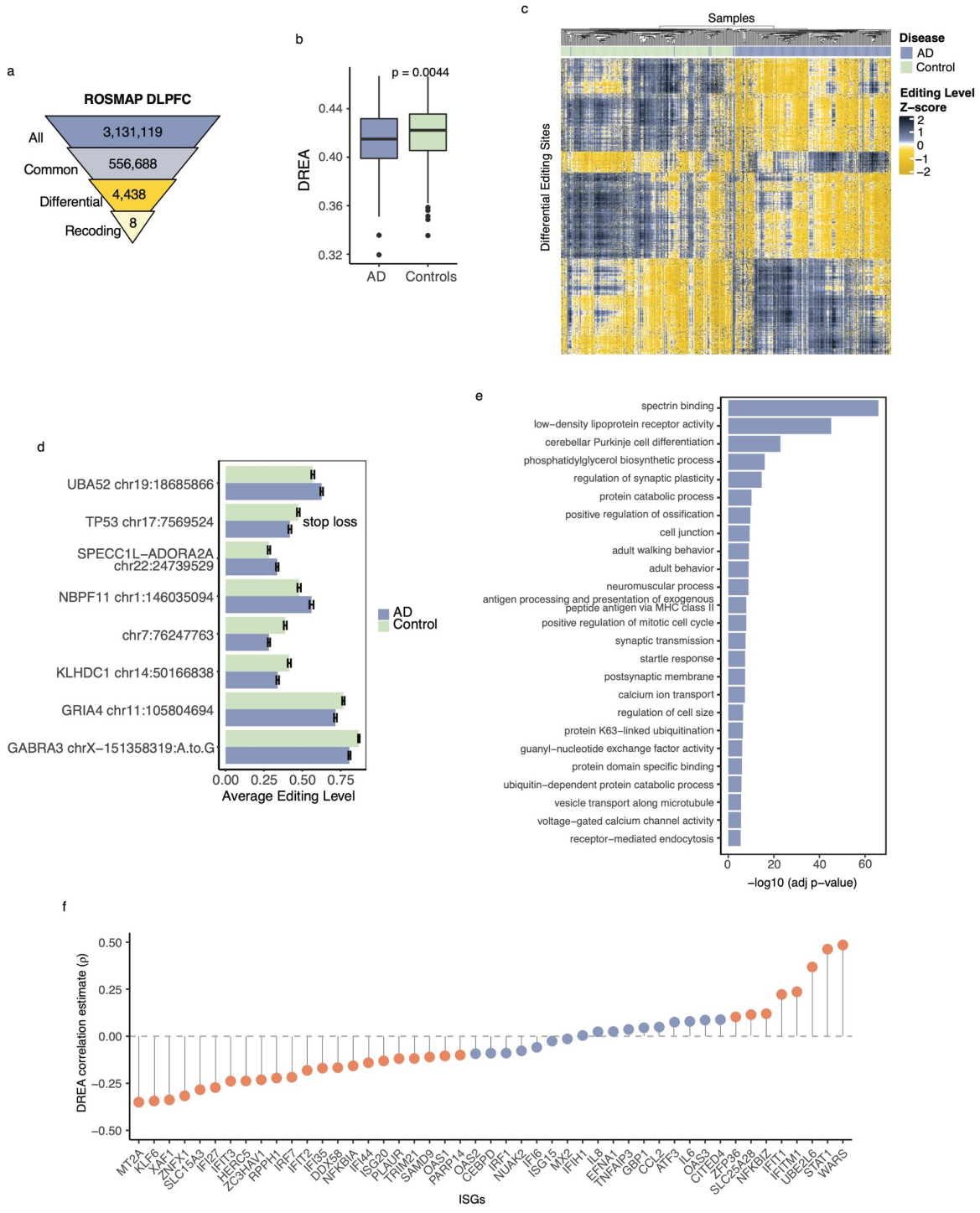
Supplementary Figure 4-2. Comparison of potential confounding variables between AD and controls in ROSMAP.

Biological covariates, technical covariates, and RNA-seq mapping metrics for AD and controls after completing the sample QC procedure (Methods). P values were calculated via the Pearson correlation and Fisher's exact test for numeric and categorical covariates, respectively.



Supplementary Figure 4-3. Characterization of de novo RNA editing sites detected in ROSMAP.

a, The total number of common editing sites detected for each ROSMAP sample, grouped by RNA-DNA difference (RDD) categories. **b,** Common sites in ALU regions and non-ALU regions, grouped by RDD categories. **c,** Distribution of differential sites in different genomic regions.



Supplementary Figure 4-4. Differential RNA editing in AD.

a, Summary of editing sites detected in each step of the RNA editing analysis. Top to bottom: All detected RDDs, common sites with non-zero editing in ≥10% of samples, sites differentially edited between AD and control, and

nonsynonymous protein recoding sites among the differential sites. **b,** Differential RNA editing average (DREA) per sample for all differential loci covered by least 5 reads. Wilcoxon rank sum p value is shown to evaluate difference between conditions. **c,** Hierarchical clustering of differential editing sites (rows) and samples (columns). Editing level z scores were calculated for each site across all samples. **d,** Editing levels in controls or AD of each differential protein recoding site. Error bars correspond to the standard error of the mean (SEM). **e,** GO enrichment results for all differentially edited genes. **f**, DREA association with ISG expression. Spearman correlation was utilized for ρ estimates and to evaluate significance of association (orange: $P < 0.05$).

CHAPTER 5 - Concluding remarks

5.1 SUMMARY

Dysregulation of RNA editing has significant consequences in brain function and immunity. While the functional roles of editing have been demonstrated for a small number of sites, the majority of RNA editing events, especially those in noncoding regions, were rarely studied for their function. To fill in this gap, we carried out global investigation of RNA editing in SCZ and AD. We demonstrated the role of editing in mitochondrial dysregulation and its potential implications in SCZ-related disease processes. Furthermore, we identified genomic loci associated with post-transcriptional RNA processing, including RNA editing, that colocalized with SCZ GWAS statistics. Finally, we applied dsRNA detection to the neurodegenerative disorder, AD, and investigated the association between dsRNA expression, editing and IFN response. Our body of work enables an improved understanding of RNA editing in brain diseases, and paves ways for the development of future novel therapies.

In Chapter 2, we identified RNA editing differences in the DLPFC of SCZ and control samples in four cohorts. We found a consistent pattern of reduced editing in the disease, and identified editing sites in various 3'UTR loci that affect mRNA abundance. Furthermore, we showed that differential editing is enriched in genes with mitochondrial functions, such as cellular respiration. Following experimental validations, we described novel functions of RNA editing sites in the mitochondrial gene, MFN1, which affect mitochondrial fission, fusion, and Cytochrome C release.

In Chapter 3, we examined edQTL, sQTL, eQTL in both EU and AA ancestries. Although shared QTL between the two ethnicities showed concordant effects on their targets, QTL unique to AA had significantly higher effect sizes than those unique to EU. These results indicate that QTL associations can be influenced by ancestral lineage, and should be considered in analyses of disease-specific editing or expression. We observed significant overlap between QTL types, although the majority of QTL were distinct. We highlighted a seedQTL located at rs146498205, which is also bound by AKAP1, an RBP known to anchor PKA to the mitochondria. Finally, we demonstrate the disease relevance of QTLs detected through colocalizations with SCZ, BPD, and MDD GWAS.

In Chapter 4, we reported dsRNAs with differential expression in AD, most of which were upregulated in the disease. We demonstrated the association between dsRNA expression and upregulated IFN response. Differentially edited loci, while demonstrating reduced editing in the disease, do not fall into differentially upregulated dsRNAs. We hypothesize that these phenomena may independently influence ISG activation. Future work is needed to decipher the mechanisms underlying the upregulation of dsRNAs and downregulation of RNA editing.

5.2 FUTURE DIRECTIONS

We demonstrated RNA editing detection in various large cohorts. Nevertheless, growing sample sizes in disease studies and rapid advances in experimental technologies continue to encourage deeper dives into disease pathologies. Our studies in no way complete the characterization of dysregulated editing in SCZ or AD, but instead, provide opportunities for further discoveries in brain disorders.

The mechanistic consequences of hypo-editing in SCZ remains to be understood. We speculate that reduced editing may lead to immune activation in the disease, which is an interesting avenue for future study. Furthermore, the dysregulation of editing in mitochondrial genes and its relationship to mitochondrial dysregulation in SCZ is unclear. Examination of functional editing sites, such as those in MFN1, represent the surface of a vast network of dysregulated molecular pathways that may contribute to disease etiology. Future investigations of noncoding and recoding sites are necessary to more specifically discern their functional consequences. For example, in vivo mouse models with differential recoding in Mfn1 may demonstrate its direct influence on mitochondrial dysregulation in brain function.

RNA-sequencing in bulk tissues is not informative of cell types or cellular compartments that undergo aberrant editing. It is likely that cell types in the brain, including neuronal subtypes, tailor their RNA editomes to flexibly adapt to their specific functionalities or cellular environments. Exploration of RNA editing in specific cellular compartments, phases, or cell types would greatly contribute to understanding the progression of editing dysregulation in brain diseases. Although accompanied by its own obstacles, RNA editing detection in single cells may allow to address many questions related to cell-type specific contributions to disease pathology.

Finally, great opportunities exist in examining and characterizing RNA editomes of various brain regions, which are known to have distinct but equally important functionalities. RNA editing has been shown to have significant functionalities in synaptic regulation and control. However, the relationship between brain regions and the effect of RNA editing on their inter-play are unknown. Such challenging questions must be

addressed to fully understand the functionality of RNA editing in the brain and develop novel therapies for relevant brain disorders.

REFERENCES

1. Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **17**, 83–96 (2016).
2. Salter, J. D., Bennett, R. P. & Smith, H. C. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem. Sci.* **41**, 578–594 (2016).
3. Wulff, B. E. & Nishikura, K. Substitutional A-to-I RNA editing. *Wiley Interdiscip. Rev. RNA* **1**, 90–101 (2010).
4. Yablonovitch, A. L., Deng, P., Jacobson, D. & Li, J. B. The evolution and adaptation of A-to-I RNA editing. *PLOS Genet.* **13**, e1007064 (2017).
5. Hung, L. Y. *et al.* An Evolutionary Landscape of A-to-I RNA Editome across Metazoan Species. *Genome Biol. Evol.* **10**, 521–537 (2018).
6. Walkley, C. R. & Li, J. B. Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. *Genome Biol.* **18**, 205 (2017).
7. Häsler, J. & Strub, K. Alu elements as regulators of gene expression. *Nucleic Acids Res.* **34**, 5491–5497 (2006).
8. Bahn, J. H. *et al.* Genomic analysis of ADAR1 binding and its involvement in multiple RNA processing pathways. *Nat. Commun. 2015 61* **6**, 1–13 (2015).
9. Bazak, L. *et al.* A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **24**, 365–76 (2014).
10. Hsiao, Y.-H. E. *et al.* RNA editing in nascent RNA affects pre-mRNA splicing. *Genome Res.* gr.231209.117 (2018). doi:10.1101/gr.231209.117
11. Nishikura, K. Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* **79**, 321 (2010).
12. La Via, L. *et al.* Modulation of dendritic AMPA receptor mRNA trafficking by RNA splicing and editing. *Nucleic Acids Res.* **41**, 617–631 (2013).
13. Irimia, M. *et al.* Evolutionarily conserved A-to-I editing increases protein stability of the alternative splicing factor Nova1. *RNA Biol.* **9**, 12–21 (2012).
14. Kortenbruck, G., Berger, E., Speckmann, E. J. & Musshoff, U. RNA editing at the Q/R site for the glutamate receptor subunits GLUR2, GLUR5, and GLUR6 in hippocampus and temporal cortex from epileptic patients. *Neurobiol. Dis.* **8**, 459–468 (2001).
15. Park, E., Jiang, Y., Hao, L., Hui, J. & Xing, Y. Genetic variation and microRNA targeting of A-to-I RNA editing fine tune human tissue transcriptomes. *Genome Biol.* **22**, 1–28 (2021).
16. Gallo, A., Vukic, D., Michalík, D., O'Connell, M. A. & Keegan, L. P. ADAR RNA editing in human disease; more to it than meets the I. *Hum. Genet.* **136**, 1265–1278 (2017).
17. Patterson, J. B. & Samuel, C. E. Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol. Cell. Biol.* **15**, 5376–5388 (1995).
18. Ward, S. V. *et al.* RNA editing enzyme adenosine deaminase is a restriction factor for controlling measles virus replication that also is required for embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 331–336 (2011).
19. Pestal, K. *et al.* Isoforms of the RNA editing enzyme ADAR1 independently

control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity* **43**, 933 (2015).
20. Vogel, O. A. *et al.* The p150 Isoform of ADAR1 Blocks Sustained RLR signaling and Apoptosis during Influenza Virus Infection. *PLOS Pathog.* **16**, e1008842 (2020).
21. Ahmad, S., Mu, X. & Hur, S. The role of rna editing in the immune response. *Methods Mol. Biol.* **2181**, 287–307 (2021).
22. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321–49 (2010).
23. Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu. Rev. Biochem.* **71**, 817 (2002).
24. Slotkin, W. & Nishikura, K. Adenosine-to-inosine RNA editing and human disease. *Genome Med.* **5**, 105 (2013).
25. Shallev, L. *et al.* Decreased A-to-I RNA editing as a source of keratinocytes' dsRNA in psoriasis. *RNA* **24**, 828–840 (2018).
26. Ramaswami, G. & Li, J. B. Identification of human RNA editing sites: A historical perspective. *Methods* **107**, 42–47 (2016).
27. 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).
28. Isaac, J. T. R., Ashby, M. & McBain, C. J. The Role of the GluR2 Subunit in AMPA Receptor Function and Synaptic Plasticity. *Neuron* **54**, 859–871 (2007).
29. Wright, A. & Vissel, B. The essential role of AMPA receptor GluA2 subunit RNA editing in the normal and diseased brain. *Front. Mol. Neurosci.* **0**, 34 (2012).
30. Shtrichman, R. *et al.* Altered A-to-I RNA Editing in Human Embryogenesis. *PLoS One* **7**, e41576 (2012).
31. Pachernegg, S., Münster, Y., Muth-Köhne, E., Fuhrmann, G. & Hollmann, M. GluA2 is rapidly edited at the Q/R site during neural differentiation in vitro. *Front. Cell. Neurosci.* **9**, 69 (2015).
32. Konen, L. M. *et al.* A new mouse line with reduced GluA2 Q/R site RNA editing exhibits loss of dendritic spines, hippocampal CA1-neuron loss, learning and memory impairments and NMDA receptor-independent seizure vulnerability. *Mol. Brain* **13**, 1–19 (2020).
33. Brusa, R. *et al.* Early-Onset Epilepsy and Postnatal Lethality Associated with an Editing-Deficient GluR-B Allele in Mice. *Science (80-.).* **270**, 1677–1680 (1995).
34. Krestel, H. E. *et al.* A Genetic Switch for Epilepsy in Adult Mice. *J. Neurosci.* **24**, 10568–10578 (2004).
35. Feldmeyer, D. *et al.* Neurological dysfunctions in mice expressing different levels of the Q/R site–unedited AMPAR subunit GluR–B. *Nat. Neurosci. 1999 21* **2**, 57–64 (1999).
36. Meier, J. C., Kankowski, S., Krestel, H. & Hetsch, F. RNA Editing-Systemic Relevance and Clue to Disease Mechanisms? *Front. Mol. Neurosci.* **9**, 124 (2016).
37. Silberberg, G., Lundin, D., Navon, R. & Öhman, M. Deregulation of the A-to-I RNA editing mechanism in psychiatric disorders. *Hum. Mol. Genet.* **21**, 311–321 (2012).

38. Sodhi, M. S., Burnet, P. W. J., Makoff, A. J., Kerwin, R. W. & Harrison, P. J. RNA editing of the 5-HT2C receptor is reduced in schizophrenia. *Mol. Psychiatry* **6**, 373–379 (2001).
39. Silberberg, G., Lundin, D., Navon, R. & Öhman, M. Deregulation of the A-to-I RNA editing mechanism in psychiatric disorders. *Hum. Mol. Genet.* **21**, 311–321 (2012).
40. Dracheva, S. *et al.* Increased serotonin 2C receptor mRNA editing: a possible risk factor for suicide. *Mol. Psychiatry* **13**, 1001–1010 (2008).
41. Tran, S. S. *et al.* Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* **22**, 25–36 (2019).
42. Wu, S., Yang, M., Kim, P. & Zhou, X. ADeditome provides the genomic landscape of A-to-I RNA editing in Alzheimer's disease. *Brief. Bioinform.* **22**, (2021).
43. Picardi, E., Horner, D. S. & Pesole, G. Single-cell transcriptomics reveals specific RNA editing signatures in the human brain. *RNA* **23**, 860–865 (2017).
44. Hwang, T. *et al.* Dynamic regulation of RNA editing in human brain development and disease. *Nat. Neurosci. 2016 198* **19**, 1093–1099 (2016).
45. Picardi, E. *et al.* Profiling RNA editing in human tissues: Towards the inosinome Atlas. *Sci. Rep.* **5**, (2015).
46. Tan, M. H. *et al.* Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**, (2017).
47. Ross, C. A., Margolis, R. L., Reading, S. A. J., Pletnikov, M. & Coyle, J. T. Neurobiology of Schizophrenia. *Neuron* **52**, 139–153 (2006).
48. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).
49. Consortium, S. W. G. of the P. G. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
50. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
51. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat. Genet.* **50**, 538–548 (2018).
52. Gavin, D. P. & Akbarian, S. Epigenetic and post-transcriptional dysregulation of gene expression in schizophrenia and related disease. *Neurobiol. Dis.* **46**, 255–262 (2012).
53. Breen, M. S. *et al.* Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nat. Neurosci.* **22**, 1402–1412 (2019).
54. Flippo, K. H. & Strack, S. An Emerging Role for Mitochondrial Dynamics in Schizophrenia. *Schizophr. Res.* **187**, 26 (2017).
55. 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).
56. Takata, A., Matsumoto, N. & Kato, T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun. 2017 81* **8**, 1–11 (2017).

57. Zaidan, H. *et al.* A-to-I RNA editing in the rat brain is age-dependent, region-specific and sensitive to environmental stress across generations. *BMC Genomics* **19**, (2018).
58. Christofi, T. & Zaravinos, A. RNA editing in the forefront of epitranscriptomics and human health. *J. Transl. Med. 2019 171* **17**, 1–15 (2019).
59. Maas, S., Kawahara, Y., Tamburro, K. M. & Nishikura, K. A-to-I RNA Editing and Human Disease. *RNA Biol.* **3**, 1 (2006).
60. Hwang, T. *et al.* Dynamic regulation of RNA editing in human brain development and disease. *Nat. Neurosci.* **19**, (2016).
61. Sebastiani, P. *et al.* RNA Editing Genes Associated with Extreme Old Age in Humans and with Lifespan in C. elegans. *PLoS One* **4**, e8210 (2009).
62. Lamers, M. M., van den Hoogen, B. G. & Haagmans, B. L. ADAR1: 'Editor-in-Chief' of Cytoplasmic Innate Immunity. *Front. Immunol.* **10**, 1763 (2019).
63. Brookmeyer, R. *et al.* National estimates of the prevalence of Alzheimer's disease in the United States. *Alzheimer's Dement.* **7**, 61–73 (2011).
64. Schultz, C., Del Tredici, K. & Braak, H. Neuropathology of Alzheimer's Disease. 21–31 (2004). doi:10.1007/978-1-59259-661-4_2
65. Kocahan, S. & Doğan, Z. Mechanisms of Alzheimer's Disease Pathogenesis and Prevention: The Brain, Neural Pathology, N-methyl-D-aspartate Receptors, Tau Protein and Other Risk Factors. *Clin. Psychopharmacol. Neurosci.* **15**, 1 (2017).
66. Weller, J. & Budson, A. Current understanding of Alzheimer's disease diagnosis and treatment. *F1000Research* **7**, (2018).
67. Budson, A. E. & Solomon, P. R. New criteria for Alzheimer disease and mild cognitive impairment: implications for the practicing clinician. *Neurologist* **18**, 356–363 (2012).
68. Jack, C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.* **14**, 535–562 (2018).
69. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimer's Dis.* **64**, S161–S189 (2018).
70. Fillenbaum, G. G. *et al.* CERAD (Consortium to Establish a Registry for Alzheimer's Disease) The first 20 years. *Alzheimers. Dement.* **4**, 96 (2008).
71. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* **112**, 389 (2006).
72. Kinney, J. W. *et al.* Inflammation as a central mechanism in Alzheimer's disease. *Alzheimer's Dement. Transl. Res. Clin. Interv.* **4**, 575 (2018).
73. Roy, E. R. *et al.* Type I interferon response drives neuroinflammation and synapse loss in Alzheimer disease. *J. Clin. Invest.* **130**, 1912–1930 (2020).
74. Schlee, M. & Hartmann, G. Discriminating self from non-self in nucleic acid sensing. *Nat. Rev. Immunol. 2016 169* **16**, 566–580 (2016).
75. Wu, S., Yang, M., Kim, P. & Zhou, X. ADeditome provides the genomic landscape of A-to-I RNA editing in Alzheimer's disease. *Brief. Bioinform.* **2021**, 1–12 (2021).
76. Gardner, O. K. *et al.* RNA editing alterations in a multi-ethnic Alzheimer disease cohort converge on immune and endocytic molecular pathways. *Hum. Mol. Genet.* **28**, 3053–3061 (2019).
77. Brümmer, A., Yang, Y., Chan, T. W. & Xiao, X. Structure-mediated modulation of

mRNA abundance by A-to-I editing. doi:10.1038/s41467-017-01459-7
78. Li, J. B. & Church, G. M. Deciphering the functions and regulation of brain-enriched A-to-I RNA editing. *Nat. Neurosci.* **16**, 1518–1522 (2013).
79. McCutcheon, R. A., Reis Marques, T. & Howes, O. D. Schizophrenia-An Overview. *JAMA psychiatry* **77**, 201–210 (2020).
80. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nat. 2022 6047906* **604**, 502–508 (2022).
81. Silberberg, G., Lundin, D., Navon, R. & Öhman, M. Deregulation of the A-to-I RNA editing mechanism in psychiatric disorders. *Hum. Mol. Genet.* **21**, 311–321 (2012).
82. Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).
83. Akbarian, S. *et al.* The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).
84. Bahn, J. H. *et al.* Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–50 (2012).
85. Lee, J.-H., Ang, J. K. & Xiao, X. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA* **19**, 725–32 (2013).
86. Zhang, Q. & Xiao, X. Genome sequence-independent identification of RNA editing sites. *Nat. Methods* **12**, (2015).
87. Porath, H. T., Carmi, S. & Levanon, E. Y. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.* **5**, (2014).
88. Picardi, E., D'Erchia, A. M., Lo Giudice, C. & Pesole, G. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **45**, D750–D757 (2017).
89. Levanon, K., Eisenberg, E., Rechavi, G. & Levanon, E. Y. Letter from the editor: adenosine-to-inosine RNA editing in Alu repeats in the human genome. *EMBO Rep.* **6**, 831–835 (2005).
90. Harjanto, D. *et al.* RNA editing generates cellular subsets with diverse sequence within populations. *Nat. Commun.* **7**, 12145 (2016).
91. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nat. 2019 5707761* **570**, 332–337 (2019).
92. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol. 2019 377* **37**, 773–782 (2019).
93. Yu, Q. & He, Z. Comprehensive investigation of temporal and autism-associated cell type composition-dependent and independent gene expression changes in human brains. *Sci. Rep.* **7**, 4121 (2017).
94. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, (2018).
95. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, (2008).
96. Yang, C. C. *et al.* ADAR1-mediated 3′ UTR editing and expression control of antiapoptosis genes fine-tunes cellular apoptosis response. *Cell Death Dis. 2017 85* **8**, e2833–e2833 (2017).
97. Siegel, D. A., Le Tonqueze, O., Biton, A., Zaitlen, N. & Erle, D. J. Massively

parallel analysis of human 3′ UTRs reveals that AU-rich element length and registration predict mRNA destabilization. *G3 Genes|Genomes|Genetics* **12**, (2022).
98. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–W220 (2010).
99. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498 (2003).
100. Youle, R. J. & Van Der Bliek, A. M. Mitochondrial fission, fusion, and stress. *Science* **337**, 1062–1065 (2012).
101. Ye, L.-Q. *et al.* The RNA editome of Macaca mulatta and functional characterization of RNA editing in mitochondria. *Sci. Bull.* **62**, 820–830 (2017).
102. Danecek, P. *et al.* High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.* **13**, 1–12 (2012).
103. Wong, R. Y. & Godwin, J. Neurotranscriptome profiles of multiple zebrafish strains. *Genomics Data* **5**, 206–209 (2015).
104. Chen, H. *et al.* Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *J. Cell Biol.* **160**, 189–200 (2003).
105. Eura, Y., Ishihara, N., Yokota, S. & Mihara, K. Two Mitofusin Proteins, Mammalian Homologues of FZO, with Distinct Functions Are Both Required for Mitochondrial Fusion. *J. Biochem.* **134**, 333–344 (2003).
106. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
107. Tondera, D. *et al.* SLP-2 is required for stress-induced mitochondrial hyperfusion. *EMBO J.* **28**, 1589–1600 (2009).
108. Zemirli, N. *et al.* Mitochondrial hyperfusion promotes NF-κB activation via the mitochondrial E3 ligase MULAN. *FEBS J.* **281**, 3095–3112 (2014).
109. Karbowski, M. & Youle, R. J. Dynamics of mitochondrial morphology in healthy cells and during apoptosis. *Cell Death Differ.* **10**, 870–880 (2003).
110. Suen, D. F., Norris, K. L. & Youle, R. J. Mitochondrial dynamics and apoptosis. *Genes Dev.* **22**, 1577–1590 (2008).
111. Ott, M., Robertson, J. D., Gogvadze, V., Zhivotovsky, B. & Orrenius, S. Cytochrome c release from mitochondria proceeds by a two-step process. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1259–1263 (2002).
112. Takuma, H., Kwak, S., Yoshizawa, T. & Kanazawa, I. Reduction of GluR2 RNA Editing, A Molecular Change that Increases Calcium Influx through AMPA Receptors, Selective in the Spinal Ventral Gray of Patients with Amyotrophic Lateral Sclerosis. (1999). doi:10.1002/1531-8249
113. Hjelm, B. E. *et al.* Evidence of Mitochondrial Dysfunction within the Complex Genetic Etiology of Schizophrenia. *Mol. Neuropsychiatry* **1**, 201 (2015).
114. Ben-Shachar, D. & Laifenfeld, D. Mitochondria, synaptic plasticity, and schizophrenia. *Int. Rev. Neurobiol.* **59**, 273–296 (2004).
115. Roberts, R. C. Postmortem studies on mitochondria in schizophrenia. *Schizophr. Res.* **187**, 17 (2017).
116. Park, C. & Park, S. K. Molecular Links between Mitochondrial Dysfunctions and

Schizophrenia. *Mol. Cells* **33**, 105 (2012).
117. Rajasekaran, A., Venkatasubramanian, G., Berk, M. & Debnath, M. Mitochondrial dysfunction in schizophrenia: pathways, mechanisms and implications. *Neurosci. Biobehav. Rev.* **48**, 10–21 (2015).
118. Millar, J. K., James, R., Christie, S. & Porteous, D. J. Disrupted in schizophrenia 1 (DISC1): subcellular targeting and induction of ring mitochondria. *Mol. Cell. Neurosci.* **30**, 477–484 (2005).
119. Devine, M. J., Norkett, R. & Kittler, J. T. DISC1 is a coordinator of intracellular trafficking to shape neuronal development and connectivity. *J. Physiol.* **594**, 5459–5469 (2016).
120. Norkett, R. *et al.* DISC1-dependent Regulation of Mitochondrial Dynamics Controls the Morphogenesis of Complex Neuronal Dendrites. *J. Biol. Chem.* **291**, 613 (2016).
121. Yamada, S., Kubo, Y., Yamazaki, D., Sekino, Y. & Kanda, Y. Chlorpyrifos inhibits neural induction via Mfn1-mediated mitochondrial dysfunction in human induced pluripotent stem cells. *Sci. Rep.* **7**, (2017).
122. Detmer, S. A. & Chan, D. C. Complementation between mouse Mfn1 and Mfn2 protects mitochondrial fusion defects caused by CMT2A disease mutations. *J. Cell Biol.* **176**, 405 (2007).
123. Sharma, S. *et al.* Mitochondrial hypoxic stress induces widespread RNA editing by APOBEC3G in natural killer cells. *Genome Biol.* **20**, 1–17 (2019).
124. Burté, F., Carelli, V., Chinnery, P. F. & Yu-Wai-Man, P. Disturbed mitochondrial dynamics and neurodegenerative disorders. *Nat. Rev. Neurol. 2014 111* **11**, 11–24 (2014).
125. Knott, A. B., Perkins, G., Schwarzenbacher, R. & Bossy-Wetzel, E. Mitochondrial fragmentation in neurodegeneration. *Nat. Rev. Neurosci. 2008 97* **9**, 505–518 (2008).
126. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol. 2019 378* **37**, 907–915 (2019).
127. Lee, J.-H., Ang, J. K. & Xiao, X. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA* **19**, 725–732 (2013).
128. Zhang, Q. & Xiao, X. Genome Sequence-Independent Identification of RNA Editing Sites. *Nat. Methods* **12**, 347 (2015).
129. Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* **43**, W589–W598 (2015).
130. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).
131. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
132. RepeatMasker Home Page. Available at: http://www.repeatmasker.org/. (Accessed: 16th November 2021)
133. CRAN - Package gplots. Available at: https://cran.r-project.org/web/packages/gplots/index.html. (Accessed: 16th November 2021)
134. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with

high-throughput sequencing data. *Bioinformatics* **31**, 166 (2015).
135. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
136. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science (80-.).* **347**, 1138–1142 (2015).
137. Zhao, W. *et al.* Massively parallel functional annotation of 3′ untranslated regions. *Nat. Biotechnol. 2014 324* **32**, 387–391 (2014).
138. Gordon, M. G. *et al.* lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. *Nat. Protoc. 2020 158* **15**, 2387–2412 (2020).
139. Ashuach, T. *et al.* MPRAnalyze: statistical framework for massively parallel reporter assays. *Genome Biol.* **20**, (2019).
140. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24 (2011).
141. Owen, M. J., Sawa, A. & Mortensen, P. B. Schizophrenia. *Lancet (London, England)* **388**, 86 (2016).
142. Haddad, P. M. & Correll, C. U. The acute efficacy of antipsychotics in schizophrenia: a review ofrecent meta-analyses. *Ther. Adv. Psychopharmacol.* **8**, 303 (2018).
143. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
144. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nat. 2022 6047906* **604**, 502–508 (2022).
145. Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771.e22 (2019).
146. Li, Q. *et al.* RNA editing underlies genetic risk of common inflammatory diseases. *Nat. 2022 6087923* **608**, 569–577 (2022).
147. Liu, H. *et al.* Functional Impact of RNA editing and ADARs on regulation of gene expression: Perspectives from deep sequencing studies. *Cell Biosci.* **4**, 1–7 (2014).
148. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
149. Choudhury, M. *et al.* Widespread RNA hypoediting in schizophrenia and its relevance to mitochondrial function. *Manuscr. Submitt.* (2022).
150. Bahn, J. H. *et al.* Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–150 (2012).
151. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet. 2017 501* **50**, 151–158 (2017).
152. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
153. Mostafavi, S. *et al.* Normalizing RNA-Sequencing Data by Modeling Hidden Covariates with Prior Knowledge. *PLoS One* **8**, e68141 (2013).
154. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).

155. Gehring, N. H. & Roignant, J. Y. Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends Genet.* **37**, 355–372 (2021).
156. Hsiao, Y.-H. E. *et al.* RNA editing in nascent RNA affects pre-mRNA splicing. *Genome Res.* gr.231209.117 (2018). doi:10.1101/gr.231209.117
157. Ali, H. S., Boshra, M. S., El Meteini, M. S., Shafei, A. E. S. & Matboli, M. lncRNA-RP11-156p1.3, novel diagnostic and therapeutic targeting via CRISPR/Cas9 editing in hepatocellular carcinoma. *Genomics* **112**, 3306–3314 (2020).
158. García-Bueno, B. *et al.* Evidence of activation of the Toll-like receptor-4 proinflammatory pathway in patients with schizophrenia. *J. Psychiatry Neurosci.* **41**, E46 (2016).
159. Yang, E.-W. *et al.* Allele-specific binding of RNA-binding proteins reveals functional genetic variants in the RNA. *Nat. Commun.* **10**, 1338 (2019).
160. Hsiao, Y. H. E. *et al.* Alternative splicing modulated by genetic variants demonstrates accelerated evolution regulated by highly conserved proteins. *Genome Res.* **26**, 440 (2016).
161. Amoah, K. *et al.* Allele-specific alternative splicing and its functional genetic variants in human tissues. *Genome Res.* **31**, 359–371 (2021).
162. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882 (2020).
163. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
164. Affaitati, A. *et al.* Essential Role of A-kinase Anchor Protein 121 for cAMP Signaling to Mitochondria *. *J. Biol. Chem.* **278**, 4286–4294 (2003).
165. Rinaldi, L. *et al.* Mitochondrial AKAP1 supports mTOR pathway and tumor growth. *Cell Death Dis. 2017 86* **8**, e2842–e2842 (2017).
166. Nucleotide BLAST: Search nucleotide databases using a nucleotide query. Available at: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. (Accessed: 15th August 2022)
167. Hook, V. *et al.* Cathepsin B in Neurodegeneration of Alzheimer's Disease, Traumatic Brain Injury, and Related Brain Disorders. *Biochim. Biophys. acta. Proteins proteomics* **1868**, 140428 (2020).
168. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, (2011).
169. Samir, P. *et al.* DDX3X acts as a live-or-die checkpoint in stressed cells by regulating NLRP3 inflammasome. *Nature* 1–5 (2019). doi:10.1038/s41586-019-1551-2
170. Mo, J. *et al.* DDX3X: structure, physiologic functions and cancer. *Mol. Cancer 2021 201* **20**, 1–20 (2021).
171. Scala, M. *et al.* Three de novo DDX3X variants associated with distinctive brain developmental abnormalities and brain tumor in intellectually disabled females. *Eur. J. Hum. Genet.* **27**, 1254 (2019).
172. Reineke, L. C. & Lloyd, R. E. The Stress Granule Protein G3BP1 Recruits Protein Kinase R To Promote Multiple Innate Immune Antiviral Responses. *J. Virol.* **89**, 2575–2589 (2015).
173. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions.

Nat. Neurosci. **22**, 343–352 (2019).
174. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
175. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
176. Nievergelt, C. M. *et al.* International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat. Commun.* **10**, (2019).
177. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
178. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genet.* **10**, e1004383 (2014).
179. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLOS Genet.* **16**, e1008720 (2020).
180. Uher, R. & Zwicker, A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry* **16**, 121–129 (2017).
181. Smoller, J. W. *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet (London, England)* **381**, 1371–1379 (2013).
182. Nuebel, E., Manganas, P. & Tokatlidis, K. Orphan proteins of unknown function in the mitochondrial intermembrane space proteome: New pathways and metabolic cross-talk. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**, 2613–2623 (2016).
183. Chen, M. *et al.* m6A RNA Degradation Products Are Catabolized by an Evolutionarily Conserved N6-Methyl-AMP Deaminase in Plant and Mammalian Cells. *Plant Cell* **30**, 1511 (2018).
184. Philips, M. A. *et al.* Lsamp is implicated in the regulation of emotional and social behavior by use of alternative promoters in the brain. *Brain Struct. Funct.* **220**, 1381 (2015).
185. Innos, J., Koido, K., Philips, M. A. & Vasar, E. Limbic system-associated membrane protein as a potential target for neuropsychiatric disorders. *Front. Pharmacol.* **4 MAR**, 32 (2013).
186. Koido, K. *et al.* Associations between LSAMP gene polymorphisms and major depressive disorder and panic disorder. *Transl. Psychiatry 2012 28* **2**, e152–e152 (2012).
187. Shi, X. *et al.* An expression signature model to predict lung adenocarcinoma-specific survival. *Cancer Manag. Res.* **10**, 3717 (2018).
188. Yan, L. *et al.* Haplotype and Cell Proliferation Analyses of Candidate Lung Cancer Susceptibility Genes on Chromosome 15q24-25.1. *Cancer Res.* **69**, 7844 (2009).
189. Ohi, K. *et al.* Genome-Wide Variants Shared Between Smoking Quantity and Schizophrenia on 15q25 Are Associated With CHRNA5 Expression in the Brain. *Schizophr. Bull.* **45**, 813 (2019).
190. Schrode, N. *et al.* Synergistic effects of common schizophrenia risk variants. *Nat. Genet. 2019 5110* **51**, 1475–1485 (2019).
191. Yang, Q. *et al.* Case reports: Three novel variants in PCCA and PCCB genes in

Chinese patients with propionic acidemia. *BMC Med. Genet.* **21**, 1–6 (2020).
192. Zhong, Y. *et al.* Discovery of novel hepatocyte eQTLs in African Americans. *PLOS Genet.* **16**, e1008662 (2020).
193. Shang, L. *et al.* Genetic Architecture of Gene Expression in European and African Americans: An eQTL Mapping Study in GENOA. *Am. J. Hum. Genet.* **106**, 496–512 (2020).
194. Flippo, K. H. *et al.* AKAP1 Protects from Cerebral Ischemic Stroke by Inhibiting Drp1-Dependent Mitochondrial Fission. *J. Neurosci.* **38**, 8233–8242 (2018).
195. Johnson-Kerner, B. *et al.* DDX3X-Related Neurodevelopmental Disorder. *GeneReviews®* (2020).
196. Jia, X. *et al.* De novo variants in genes regulating stress granule assembly associate with neurodevelopmental disorders. *Sci. Adv.* **8**, 17 (2022).
197. Jarskog, L. F., Glantz, L. A., Gilmore, J. H. & Lieberman, J. A. Apoptotic mechanisms in the pathophysiology of schizophrenia. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **29**, 846–858 (2005).
198. Müller, N. & Schwarz, M. J. Immune System and Schizophrenia. *Curr. Immunol. Rev.* **6**, 213 (2010).
199. Josse, J. & Husson, F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. *J. Stat. Softw.* **70**, 1–31 (2016).
200. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353 (2012).
201. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).
202. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31**, 418–420 (2015).
203. Psychiatric Genomics Consortium | Psychiatric Genomics Consortium. Available at: https://www.med.unc.edu/pgc/. (Accessed: 15th August 2022)
204. Liu, B., Gloudemans, M. J., Rao, A. S., Ingelsson, E. & Montgomery, S. B. Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.* **51**, 768–769 (2019).
205. Gantier, M. P. & Williams, B. R. G. The response of mammalian cells to double-stranded RNA. *Cytokine Growth Factor Rev.* **18**, 363 (2007).
206. Brisse, M. & Ly, H. Comparative structure and function analysis of the RIG-I-like receptors: RIG-I and MDA5. *Front. Immunol.* **10**, 1586 (2019).
207. Whipple, J. M. *et al.* Genome-wide profiling of the C. elegans dsRNAome. *RNA* **21**, 786–800 (2015).
208. Blango, M. G. & Bass, B. L. Identification of the long, edited dsRNAome of LPS-stimulated immune cells. *Genome Res.* **26**, 852–862 (2016).
209. Reich, D. P., Tyc, K. M. & Bass, B. L. C. elegans ADARs antagonize silencing of cellular dsRNAs by the antiviral RNAi pathway. *Genes Dev.* **32**, 271–282 (2018).
210. Liddicoat, B. J. *et al.* RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115–1120 (2015).
211. Hartner, J. C., Walkley, C. R., Lu, J. & Orkin, S. H. ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. *Nat. Immunol. 2008 101* **10**, 109–115 (2008).

212. Mannion, N. M. *et al.* The RNA-Editing Enzyme ADAR1 Controls Innate Immune Responses to RNA. *Cell Rep.* **9**, (2014).
213. Reich, D. P. & Bass, B. L. Mapping the dsRNA World. *Cold Spring Harb. Perspect. Biol.* **11**, a035352 (2019).
214. Ahmad, S. *et al.* Breaching Self-Tolerance to Alu Duplex RNA Underlies MDA5-Mediated Inflammation. *Cell* **172**, 797-810.e13 (2018).
215. Faghihi, M. A. *et al.* Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nat. Med. 2008 147* **14**, 723–730 (2008).
216. Nakahama, T. *et al.* Mutations in the adenosine deaminase ADAR1 that prevent endogenous Z-RNA binding induce Aicardi-Goutières-syndrome-like encephalopathy. *Immunity* **54**, 1976-1988.e7 (2021).
217. Rodriguez, S. *et al.* Genome-Encoded Cytoplasmic Double-Stranded RNAs, Found in C9ORF72 ALS-FTD Brain, Propagate Neuronal Loss. *Sci. Transl. Med.* **13**, (2021).
218. Hideyama, T. *et al.* Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. *Neurobiol. Dis.* **45**, 1121–1128 (2012).
219. Schöll, M. *et al.* PET Imaging of Tau Deposition in the Aging Human Brain. *Neuron* **89**, 971–982 (2016).
220. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science (80-.).* **297**, 353–356 (2002).
221. Mastrangelo, M. A., Sudol, K. L., Narrow, W. C. & Bowers, W. J. Interferon-γ Differentially Affects Alzheimer's Disease Pathologies and Induces Neurogenesis in Triple Transgenic-AD Mice. *Am. J. Pathol.* **175**, 2076 (2009).
222. Khermesh, K. *et al.* Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA* **22**, 290–302 (2016).
223. Gardner, O. K. *et al.* RNA editing alterations in a multi-ethnic Alzheimer disease cohort converge on immune and endocytic molecular pathways. *Hum. Mol. Genet.* **28**, 3053 (2019).
224. De Jager, P. L. *et al.* A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data 2018 51* **5**, 1–13 (2018).
225. Greenwood, A. K. *et al.* The AD Knowledge Portal: A Repository for Multi-Omic Data on Alzheimer's Disease and Aging. *Curr. Protoc. Hum. Genet.* **108**, (2020).
226. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, 1–17 (2014).
227. Hur, S. Double-Stranded RNA Sensors and Modulators in Innate Immunity Keywords dsRNA, RIG-I-like receptor, protein kinase R, oligoadenylate synthase, dsRNA-dependent adenosine deaminase, RNA interference. (2019). doi:10.1146/annurev-immunol-042718
228. Scadden, A. D. J. & Smith, C. W. J. A ribonuclease specific for inosine-containing RNA: a potential role in antiviral defence? *EMBO J.* **16**, 2140 (1997).
229. Licastro, F. *et al.* Increased plasma levels of interleukin-1, interleukin-6 and α-1-antichymotrypsin in patients with Alzheimer's disease: peripheral inflammation or signals from the brain? *J. Neuroimmunol.* **103**, 97–102 (2000).

230. Remarque, E. J. *et al.* Patients with Alzheimer's disease display a pro-inflammatory phenotype. *Exp. Gerontol.* **36**, 171–176 (2001).
231. Dube, U. *et al.* An atlas of cortical circular RNA expression in Alzheimer disease brains demonstrates clinical and pathological associations. *Nat. Neurosci.* **22**, 1903–1912 (2019).
232. Readhead, B. *et al.* Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. *Neuron* **99**, 64-82.e7 (2018).
233. Zhou, H. *et al.* Genome-wide association study identifies glutamate ionotropic receptor GRIA4 as a risk gene for comorbid nicotine dependence and major depression. *Transl. Psychiatry* **8**, (2018).
234. Roth, S. H., Levanon, E. Y. & Eisenberg, E. Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat. Methods* (2019). doi:10.1038/s41592-019-0610-9
235. Chitrakar, A. *et al.* Introns encode dsRNAs undetected by RIG-I/MDA5/interferons and sensed via RNase L. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2102134118 (2021).
236. Li, H. D. *et al.* Integrative functional genomic analysis of intron retention in human and mouse brain with Alzheimer's disease. *Alzheimer's Dement.* **17**, 984–1004 (2021).
237. Samuel, C. E. ADARs, Viruses and Innate Immunity. *Curr. Top. Microbiol. Immunol.* **353**, 163–195 (2012).
238. Lange, S. J. *et al.* Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res.* **40**, 5215 (2012).
239. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–31 (2003).
240. Kato, H. *et al.* Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid–inducible gene-I and melanoma differentiation–associated gene 5. *J. Exp. Med.* **205**, 1601 (2008).
241. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15 (2013).
242. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
243. Christensen, R. H. B. Regression Models for Ordinal Data [R package ordinal version 2015.6-28].
244. Make Complex Heatmaps • ComplexHeatmap. Available at: https://jokergoo.github.io/ComplexHeatmap/. (Accessed: 11th September 2022)