Published in final edited form as:
Nat Neurosci. 2019 January ; 22(1): 25–36. doi:10.1038/s41593-018-0287-x.

Widespread RNA editing dysregulation in brains from autistic individuals

Stephen S. Tran[1,2], **Hyun-Ik Jun**[2], **Jae Hoon Bahn**[2], **Adel Azghadi**[2], **Gokul Ramaswami**[3], **Eric L. Van Nostrand**[4,5,6], **Thai B. Nguyen**[4,5,6], **Yun-Hua E. Hsiao**[7], **Changhoon Lee**[3], **Gabriel A. Pratt**[4,5,6,8], **Verónica Martínez-Cerdeño**[9], **Randi J. Hagerman**[10], **Gene W. Yeo**[4,5,6,8], **Daniel H. Geschwind**[3,11,12,*], and **Xinshu Xiao**[1,2,13,14,*]

[1]Bioinformatics Interdepartmental Program, UCLA, Los Angeles, California, CA, 90095, USA

[2]Department of Integrative Biology and Physiology, UCLA, Los Angeles, California, CA, 90095, USA

[3]Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, UCLA, Los Angeles, California, CA, 90095, USA

[4]Department of Cellular and Molecular Medicine, UCSD, La Jolla, CA, 92093, USA

[5]Stem Cell Program, UCSD, La Jolla, CA, 92093, USA

[6]Institute for Genomic Medicine, UCSD, La Jolla, CA, 92093, USA

[7]Department of Bioengineering, UCLA, Los Angeles, California, CA, 90095, USA

[8]Bioinformatics and Systems Biology Graduate Program, UCSD, La Jolla, CA, 92093, USA

---

[*]Correspondence: dhg@mednet.ucla.edu, gxxiao@ucla.edu.

Author contributions
S.S.T.carried out data analyses, with input from G.R.. H.I.J., J.H.B. and A.A. performed molecular biology experiments. E.L.V.N., T.B.N., G.A.P. and G.W.Y carried out eCLIP experiments and data processing. Y.H.E.H. contributed to data visualization. C.L. carried out ASD RNA-Seq data generation. V.M.C. and R.J.H. provided postmortem Fragile X and control samples. S.S.T, D.H.G. and X.X. designed the study, interpreted the results and wrote the manuscript.

Accession Codes
Data generated by this study are available at the Gene Expression omnibus with IDs: GSE107895, GSE107867, GSE117776.

Competing Interests Statement
G.W.Y, is a co-founder of Locana and Eclipse Bioinnovations and member of the scientific advisory boards of Locana, Eclipse Bioinnovations and Aquinnah Pharmaceuticals. E.V.N, is a co-founder and member of the scientific advisory board of Eclipse BioInnovations. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

Code availability
R code for the analysis of RNA editing of ASD against controls is available as Supplementary Software. It is also available in GitHub (where updates will be released): https://github.com/gxiaolab/RNA_editing_in_ASD.Nat_neuroscience.Tran_et_al

Data availability
eCLIP-seq data on FMRP and FXR1P from postmortem human brain have been deposited to GEO with accession GSE107895. RNA-seq data of Fragile X patients, carriers, and controls have been deposited to GEO with accession GSE107867 (NeuroBiobank dataset) and GSE117776 (UC Davis FXTAS dataset). Fastq files of RNA-Seq from the idiopathic ASD, dup15q, and control brains were obtained from our previous study[9] and are available in the PsychENCODE website (https://www.synapse.org//#!Synapse:syn4921369/wiki/235539). Fastq files of RNA-seq data from the replicate ASD and control cohort are available on GEO (accession GSE51264 / GSE59288).

[9]Department of Pathology and Laboratory Medicine, UC Davis School of Medicine, Sacramento, CA, 95817, USA

[10]The MIND Institute, Department of Pediatrics, UC Davis School of Medicine, Sacramento, CA, 95817, USA

[11]Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, UCLA, CA, 90095, USA.

[12]Department of Human Genetics, David Geffen School of Medicine, UCLA, California 90095, USA.

[13]Molecular Biology Institute, UCLA, Los Angeles, California, CA, 90095, USA

[14]Institute for Quantitative and Computational Biology, UCLA, Los Angeles, California, CA, 90095, USA

Abstract

Transcriptomic analyses of postmortem brains have begun to elucidate the molecular abnormalities in autism spectrum disorder (ASD). However, a crucial pathway involved in synaptic development, RNA editing, has not yet been studied on a genome-wide scale. Here, we profiled global patterns of adenosine-to-inosine (A-to-I) editing in a large cohort of post-mortem ASD brains. We observed a global bias for hypoediting in ASD brains, which was shared across brain regions and involved many synaptic genes. We show that the Fragile X proteins, FMRP and FXR1P, interact with ADAR proteins and modulate A-to-I editing. Furthermore, we observed convergent patterns of RNA editing alterations in ASD and Fragile X syndrome, establishing this as a molecular link between these related diseases. Our findings, which are corroborated across multiple datasets, including dup15q cases associated with intellectual disability, highlight RNA editing dysregulation in ASD and reveal novel mechanisms underlying this disorder.

Introduction

Autism spectrum disorder (ASD) is characterized by a developmental deficit in social communication accompanied by repetitive and restrictive interests[1], with a strong neuropathology implicating glutamatergic[2] and serotonergic[3] circuits, aberrant structural development in multiple brain regions[4], excitatory and inhibitory imbalance[5], and abnormal synaptogenesis[6]. The genetic etiology of ASD remains incompletely understood and shows substantial heterogeneity[7]. Nevertheless, recent studies, leveraging the increasing availability of postmortem samples, have revealed shared patterns of transcriptome dysregulation affecting neuronal and glial coding and non-coding gene expression[8,9], neuronal splicing including microexons[10], and microRNA targeting[11] across approximately 2/3 of ASD patients. These studies highlight down-regulation of activity-dependent genes in neurons and up-regulation of astrocyte and microglial genes as key points of convergence in ASD pathology.

Another major RNA processing mechanism is RNA editing, which refers to the alteration of RNA sequences through insertion, deletion or substitution of nucleotides. Catalyzed by the ADAR family of enzymes, adenosine-to-inosine (A-to-I) editing is the most prevalent type of RNA editing in humans, affecting the majority of human genes[12]. As inosines in RNA are recognized as guanosines by cellular machinery, A-to-I editing can alter gene expression in different ways, for example, through amino acid substitutions, modulation of RNA stability, alteration of alternative splicing, and modifications of regulatory RNAs or *cis*-regulatory motifs[12,13].

RNA editing plays important roles in neurodevelopment and maintenance of normal neuronal function[13]. Indeed, a number of A-to-I editing sites alone are imperative in modulating excitatory responses and permeability of ionic channels and other neuronal signaling functions[13]. Not surprisingly, aberrant RNA editing has been reported in several neurological disorders, such as schizophrenia, bipolar disorder, amyotrophic lateral sclerosis,[14] and Alzheimer's disease.[15] In ASD, a previous study analyzed a few known RNA editing sites in synaptic genes and reported altered editing patterns in a small cohort of ASD cerebella[16]. Yet, it remains unaddressed if global patterns of RNA editing may contribute to the neuropathology of ASD, a question that requires larger patient cohorts and multiple implicated brain regions. In addition, the regulatory mechanisms of aberrant editing in neurological disorders including ASD remain largely unknown.

Here we report global patterns of dysregulated RNA editing across the largest cohort of ASD brain samples to date, spanning multiple brain regions. We identified a core set of down-regulated RNA editing sites, enriched in genes of glutamatergic and synaptic pathways and ASD susceptibility genes. Multiple lines of evidence associate a distinct set of these hypoedited sites with Fragile X proteins: FMRP and FXR1P. Through transcriptome-wide protein-RNA binding analyses and detailed molecular assays, we show that FMRP and FXR1P interact with ADAR and modulate A-to-I editing. It is known that mutations in FMRP lead to the Fragile X syndrome, a disease with high comorbidity with ASD[7]. Indeed, we observed convergent dysregulated patterns of RNA editing in Fragile X and ASD patients, which is consistent with the findings that genes harboring ASD risk mutations are enriched in FMRP targets[17,18]. Overall, we provide global insights regarding RNA editing in ASD pathogenesis and elucidate a regulatory function of Fragile X proteins in RNA editing that additionally serves as a molecular link between ASD and Fragile X Syndrome.

Results

RNA editing analysis of ASD postmortem brain samples

From 69 unique post-mortem subjects, we obtained rRNA-depleted total RNA-Seq (50 base paired-end, non-strand-specific) from three brain regions implicated in ASD-susceptibility: frontal cortex, temporal cortex, and cerebellum (Supplementary Table 1). In total, there were 29, 30, and 31 ASD samples, and 33, 27, and 29 control samples from frontal cortex, temporal cortex, and cerebellum, respectively, with 45 subjects in common across 3 brain regions and 20 subjects in common across 2 regions (Supplementary Fig. 1). These datasets were generated as part of our transcriptomic study of ASD brain[9]. Overall, the ASD and control groups did not have significant differences in variables that might confound RNA editing analysis (e.g. age, gender, etc.) (Supplementary Fig. 2). Each brain sample was sequenced to an average of 70 million raw read pairs and averaged 55 million uniquely mapped pairs (Supplementary Fig. 3)[9].

We applied our previously developed methods to identify RNA editing sites using the RNA-Seq data[19], and implemented additional steps to capture editing sites located in "hyperedited" regions, which were likely missed by regular methods[20] (Methods). Combining these approaches, we identified a total of 98,477, 97,994, and 134,085 predicted RNA editing sites from frontal cortex, temporal cortex, and cerebellum, respectively. As expected, the number of predicted RNA editing sites per sample correlated with read coverage approximately (Supplementary Fig. 4).

On average, >95% predicted RNA editing sites were A-to-G and T-to-C editing types per sample, and the remaining 5% mainly consisted of C-to-T and G-to-A types, consistent with canonical A-to-I and C-to-U editing reflected in non-strand-specific RNA-Seq data (Fig. 1a, Supplementary Fig. 5). Notably, most (84%) of the A-to-I editing sites are listed in the REDIportal database[21] (Supplementary Fig. 5). The majority of RNA editing sites were located in Alu sequences[21] (Supplementary Fig. 5) and in intronic regions[21] (Supplementary Fig. 5), and the sequence context of A-to-G sites was consistent with the typical sequence signature known for ADAR substrates[22] (Supplementary Fig. 5). Examination of correlation between ADAR expression levels and various partitions of editing sites (Alu, non-Alu repetitive, non-repetitive regions) showed overall positive correlation with ADAR1 and ADAR2 across the editome and weakly negative or no correlation with ADAR3 (Supplementary Fig. 6). These findings are consistent with known properties of RNA editing established in the literature[23], and altogether, strongly support the validity of our predicted A-to-I editing sites.

The frontal and temporal cortex shared more than 70% of their sites, while the two cortical regions and cerebellum shared 50–55% (Supplementary Fig. 5). Furthermore, the editing levels of common editing sites between two brain regions were highly consistent (correlation coefficient 0.96 between cortices, and 0.89 to 0.90 between cortex and cerebellum, Supplementary Fig. 5). Thus, the three brain regions demonstrated similarities and differences in RNA editomes, with the cortices having more similarities in RNA editomes than with cerebellum, likely reflecting the substantial differences in cellular composition and physiology between these two regions[24].

Reduction of RNA editing in ASD frontal cortex

Given the observed difference in RNA editing between brain regions, we first focused on analysis of RNA editing dysregulation in frontal cortex, a region with strong transcriptomic alterations in ASD[8,9]. We identified a total of 3,314 differential editing sites in ASD ($p < 0.05$, and editing level difference ≥ 5% or editing prevalence difference ≥ 5%, see Methods and Supplementary Table 2), which were robust to the choice of statistical models and parameters (Methods, Figs. S7–9). For each individual, 2.6–10.5% of all editing sites were identified as differential (Fig. 1b). Strikingly, the differentially edited sites showed a bias of hypoediting in ASD samples (Fig. 1c); the number of down-regulated RNA editing sites in ASD far outnumbered those that were upregulated ($p = 1.3e\text{-}59$, Chi-squared test, Fig. 1c).

Across potentially confounding biological and technical variables, diagnosis (i.e., ASD or control) was the only variable with significant association (Supplementary Fig. 10), allowing differential editing sites to substantially separate the two groups of subjects (Fig. 1d). Also, genes harboring the differential editing sites had minimal gene expression differences between ASD and control groups (Supplementary Fig. 10), evidencing that differential editing was unlikely secondary to differential gene expression.

We utilized Sanger sequencing to confirm the observed editing differences of 8 sites (Supplementary Table 3), covering an expansive range of editing level differences (Fig. 1e). Each editing site was tested in eight postmortem frontal cortex samples (4 ASD, 4 controls), selected based on sample availability (Supplementary Table 1b). The editing differences calculated from RNA-Seq strongly correlated with those from Sanger sequencing (Fig. 1e, $R^2 = 0.75$), validating the accuracy of our editing level quantification.

The set of genes harboring at least one differential editing site in frontal cortex (total of 1,189) exhibited significant gene ontology (GO) enrichment for categories including ionotropic glutamate receptor activity, glutamate gated ion channel activity, and synaptic transmission (Fig. 1f). Consistently, genes (e.g., *KCNIP4*, *PCDH9*, *RBFOX1*, and *CNTNAP2*) with the largest number of differential editing sites (both before or after correction for gene length, Supplementary Fig. 11) were involved in the above functional categories, and a number of genes with differential editing were also known ASD susceptibility genes[25] (Supplementary Fig. 11). For a relatively small number of genes (Supplementary Table 4), such as *KCND2* and *GRIK2*, that harbored differential editing sites associated with their gene expression, we observed strong enrichment in synaptic functions, including presynaptic and postsynaptic membrane, synaptic transmission, cell junction, dendrites, and similar categories (Supplementary Fig. 11). Lastly, we observed that differential editing sites were significantly enriched in clusters of editing sites that abruptly increase between fetal and infant stages of cortical development[26] (Supplementary Fig. 12). Together, these results indicate that RNA editing could contribute towards aberrant synaptic formation in ASD.

Replication of reduced RNA editing in an independent cohort of ASD frontal cortex

For replication, we analyzed an independent cohort of ASD patients[27]. After balancing technical covariates, we analyzed RNA-Seq data from frontal cortex of 22 ASD and 23 controls (Supplementary Fig. 13, Supplementary Table 1c). This data set had single-end reads from polyA primed libraries and low sequencing depth (< 12 million total reads per sample, which led to slight 5’ to 3’ bias), constricting sufficient coverage to only 4952 editing sites. We, nevertheless, identified differential editing in 185 sites, with 65% exhibiting reduced editing in ASD (Supplementary Fig. 14, Chi-squared test $p = 0.0085$), thus reproducing the hypoediting pattern of our main dataset. Differential editing sites in the replication dataset were likewise enriched in genes involved in synapse and cell junction (Supplementary Fig. 14), and the levels of differential editing significantly correlated with those in our study (Supplementary Fig. 14). Replication of the editing landscape using data from a different cohort collected by a different lab strongly supports the validity of our observed ASD editing profiles.

Global analysis of potential regulators of hypoediting in ASD

To elucidate the regulatory mechanisms of hypoediting in ASD brains, we examined the mRNA and protein expression levels of the *ADAR* genes but did not observe significant differences of *ADAR1* and *ADARB1 (ADAR2)* expression in frontal cortex (Fig. 2a-c). Although *ADARB2* (*ADAR3*) protein was undetectable in the brain samples (Supplementary Fig. 15), its mRNA was slightly downregulated in ASD (Fig. 2a), which, as an RNA editing inhibitor[23], cannot explain the observed hypoediting in ASD. The *ADAR* genes did not exhibit differential splicing in these samples, as determined previously[9], and have no reported rare or common variants associated with ASD.

Given the absence of explanatory variation by *ADARs*, we hypothesized other trans-regulators must causally contribute. Given the large-scale editome profiles in this study, if a prevailing mechanism exists for hypoediting in ASD, then a significant number of editing sites should demonstrate correlated variation across the subjects. We applied weighted gene co-expression network analysis (WGCNA)[28] to search for highly correlated clusters of editing sites (i.e., modules) (Methods).

Remarkably, we identified a module enriched in editing sites that had significant association with diagnosis (Fig. 2d, Supplementary Table 5) and enrichment with differential editing sites between ASD and controls in frontal cortex of this study (Fig. 2e), and those from the replication cohort (Supplementary Fig. 14). Correlation between the module "eigengene" (i.e., eigen-editing site) and expression of potential *trans*-regulators (Supplementary Fig. 15) identified strong association between the turquoise module and Fragile X-relevant genes (*FMR1* and *FXR1*) (Fig. 2d). *FMR1* demonstrated positive (i.e. enhancing) correlation with editing changes, while *FXR1* displayed negative (i.e. inhibitory) correlation. This module is significantly enriched with genes related to synaptic ontology (Fig. 2f), consistent with a primary known function of FMRP in localization and maintenance of synapses[29], and previous reports showing enrichment of FMRP binding targets in ASD risk genes[17,18].

Interaction between Fragile X proteins and ADARs

To experimentally inspect the involvement of Fragile X proteins in RNA editing regulation, we first conducted subcellular fractionation experiment followed by Western blot and reciprocal co-immunoprecipitation (co-IP) experiments in HeLa cells to determine the localization and protein interactions of Fragile X proteins and ADARs. Consistent with previous literature, the ADAR proteins were enriched in the nuclear fraction[12], while FMRP and FXR1P were detected substantially in the cytoplasmic fraction[30] (Fig. 3a). Interestingly, FMRP and FXR1P were also highly detectable in nucleus, which was corroborated using immunofluorescence experiments (Supplementary Fig. 16). Subcellular distribution of ADAR proteins remained unchanged upon FMRP or FXR1P knockdown (Fig. 3a). Reciprocal co-IP experiments showed that FMRP interacts with both ADAR1 and ADAR2 in an RNA-independent manner (Fig. 3b), while FXR1P interacted with ADAR1 but not with ADAR2. Additionally, we observed interaction between FMRP and FXR1P (consistent with previous literature)[31], but not between ADAR1 and ADAR2.

FMRP and FXR1P binding relative to dysregulated editing sites

Next, we captured the transcriptome-wide binding patterns of FMRP and FXR1P to RNA transcripts using enhanced UV crosslinking and immunoprecipitation (eCLIP)[32]. Data from two eCLIP experiments and an input control experiment were obtained for each protein using postmortem frontal cortex from control subjects (Methods) (Supplementary Fig. 17).

We first confirmed the quality of our eCLIP experiments. eCLIP peaks identified in each replicate (Methods, Supplementary Table 6) demonstrated highly correlated read abundance (Supplementary Fig. 17), prompting us to combine peaks from the replicate experiments to maximize the sensitivity of peak detection. The binding sites of both proteins were predominantly distributed in genic 3’ UTRs, introns and exons (Supplementary Fig. 17), consistent with previous literature[30,33]. Sequence motif analyses identified ACUG as the most enriched motif among the FMRP eCLIP peaks (Supplementary Fig. 17), which matches a FMRP binding motif previously reported[33], and CAUGC in FXR1P (Supplementary Fig. 17), which is consistent with a previous report that FXR1P tends to associate with AU-rich elements[34].

Next, we examined the proximity of FMRP and FXR1P binding peaks relative to dysregulated editing sites in ASD frontal cortex. Remarkably, the FMRP and FXR1P eCLIP peaks were significantly enriched around editing sites in the turquoise module (Fig. 3c, Methods), a finding that replicated in the FMRP eCLIP data generated from K562 cells by ENCODE[35] (Supplementary Fig. 18), but, importantly, not for proteins lacking evidence for RNA editing regulation (Supplementary Fig. 18). Additionally, FMRP and FXR1P eCLIP target genes significantly overlapped with genes harboring differential editing sites or sites in the turquoise module (Supplementary Fig. 19). These results suggest that FMRP and FXR1P proteins may regulate RNA editing directly in ASD.

FMRP directly modulates RNA editing

To investigate whether FMRP directly affects RNA editing, we conducted a series of minigene reporter assays (Supplementary Fig. 21, Methods) on two example editing sites in HeLa cells (Supplementary Table 3). These editing sites, located in the 3’ UTRs of the *TEAD1* and *EEF2K* genes, were chosen due to close proximity with putative FMRP binding motifs (Supplementary Fig. 20). The *TEAD1* and *EEF2K* editing sites are likely site-specific editing sites, since no other sites were observed in their immediate neighborhood.

Knockdown of *FMR1* and *ADAR2* caused significant reduction of editing at the *TEAD1* editing site (Fig. 3d). Similarly, knockdown of *FMR1* caused significant reduction of *EEF2K* editing level (Fig. 3e) and a trend of reduction upon *ADAR1* knockdown (p = 0.06). *EEF2K* is also endogenously edited in HeLa cells, and responded to *FMR1* and *ADAR1* knockdown significantly, concordant with the minigene assays (Supplementary Fig. 21). These data are consistent with our observation that FMRP multifariously interacts with both ADAR1 and ADAR2 proteins and corroborates the positive association of turquoise eigen-editing site with *FMR1* expression levels (Fig. 2d).

Next, we introduced mutations to the FMRP binding motifs in the minigenes in order to weaken the protein-RNA interaction (Supplementary Fig. 20). Loss of these FMRP binding sites caused significant reduction in RNA editing (Fig. 3f, g), importantly, without changing the predicted double-stranded RNA (dsRNA) structures (Supplementary Fig. 20). Our results suggest that FMRP directly regulates editing of these two site-specific sites through mediated interaction between ADAR and the RNA.

FXR1P regulates hyperedited sites

In contrast to site-specific editing, another class of editing sites consists of hyperedited sites that tend to cluster together[20]. We conducted minigene experiments on three genes (*CNTNAP4*, *NLGN1*, and *TENM2*) that all had manifold editing sites within long double-stranded intronic regions (Supplementary Fig. 22, Supplementary Table 3), two of which (*CNTNAP4* and *NLGN1*) are ASD risk genes. Consistent with its known role in hyperedited RNA editing[20], *ADAR1* knockdown caused reduction in all the detectable editing sites (Fig. 3h), though interestingly *ADAR2* knockdown did too to a lesser degree. Remarkably, the hyperedited sites in these genes showed increased editing levels in *FXR1* (but not in *FMR1*) knockdown cells, which was again consistent with the WGCNA results that showed negative correlation between *FXR1* expression and RNA editing (Fig. 2d). RNA immunoprecipitation experiments supported that FXR1P binds to the regions harboring the editing sites in these target genes (Supplementary Fig. 23). Additionally, mutations in predicted FXR1 binding motifs induced higher editing levels at a majority of sites in two of the three minigenes (Fig. 3i, Supplementary Fig. 23). These results potentially indicate direct inhibitory regulation of hyperediting sites by FXR1P through mediated interaction between ADAR1 and RNA.

Concomitant regulation of RNA editing by FMRP and FXR1P

To further substantiate the above findings, we validated 6 more differential editing sites in two neuroblastoma cell lines (Supplementary Table 3). These candidate sites were chosen based on their propinquity to FMRP or FXR1P eCLIP sites and their nominal correlation with the turquoise module, *FMR1*, or *FXR1* gene expression. As expected, *ADAR1* and *ADAR2* shRNA knockdown reduced editing at all editing sites (Supplementary Fig. 24). Strikingly, *FMR1* shRNA knockdown caused significant reduction of editing of all sites, while *FXR1* knockdown caused significant augmentation of editing in 10 of the 12 sites (Fig. 3j, S24). These results were reproducible between the two cell lines, further substantiating the inhibitory role of FXR1P and enhancing role of FMRP in editing regulation, and demonstrating concomitant regulation of RNA editing by these proteins at some editing sites. Together, our experimental results clearly validate that FMRP and FXR1P are important regulators of RNA editing.

Convergent RNA editing alterations between ASD and Fragile X patients

Loss of FMRP manifests in Fragile X syndrome, the most prevalent monogenic cause of ASD (1–2% of all ASD)[7,36] in which approximately 50% of patients have co-diagnoses or features of ASD[37]. To investigate a possible role for RNA editing contributing to shared molecular pathologies, we generated RNA-Seq data from the frontal cortex of four patients with Fragile X syndrome and four Fragile X carriers or controls (Supplementary Fig. 25). The samples were obtained and separately analyzed from two brain banks. Western blot confirmed that FMRP expression was absent or reduced in the Fragile X samples relative to carriers or controls, and the expression levels of *ADAR1* and *ADAR2* were similar between the two groups (Supplementary Fig. 25).

Strikingly, differential editing sites identified in the Fragile X dataset (Supplementary Fig. 25, Methods) showed the same trends as those from ASD: they demonstrated a predominant trend of hypoediting in Fragile X patients and strong enrichment in genes related to synaptic transmission, cellular junctions, and ionic transmission (Fig. 4a,b, Supplementary Table 7), and were also significantly enriched around FMRP and FXR1P eCLIP peaks (Fig. 4c). Moreover, a statistically significant overlap was observed between the differentially edited genes in Fragile X patients and those in the turquoise module identified from ASD frontal cortex (Fig. 4d), the module that is correlated with *FMR1* expression (Fig. 2d). In addition, a significant overlap exists between the differential editing sites in Fragile X patients and editing sites in the turquoise module of ASD for data from one of the two brain banks (Supplementary Fig. 26). Overall, these results again support our hypothesis that the turquoise module encapsulates a subset of dysregulated editing sites in ASD that are under regulation by FMRP.

Altogether, the analysis of editing profiles in Fragile X patient brain provides a strong independent line of evidence showing convergence of dysregulated RNA editing between Fragile X syndrome and ASD through a common mechanism involving FMRP regulation of RNA editing.

Consistent hypoediting patterns observed for different brain regions of ASD patients

Here we investigated whether other brain regions share similar editing patterns with the frontal cortex. In temporal cortex and cerebellum, we also observed global down-regulation of RNA editing and enrichment in synapses, cellular junctions, and ionic channels (Fig. 5a and Supplementary Fig. 27). Overall, differential editing sites shared between brain regions showed significant correlation in levels of dysregulation (Fig. 5b). Likewise, WGCNA, performed on the editing sites identified in temporal cortex and cerebellum, identified downregulated modules (colored turquoise by WGCNA convention) strongly associated with ASD in these brain regions respectively (Fig. 5c, Supplementary Table 5). The turquoise modules of the three brain regions shared many editing sites (Fig. 5d). Overall, these results demonstrate that the global patterns of dysregulated editing are common across implicated brain regions in ASD.

A small set of 65 and 66 genes were, however, exclusively differentially edited in cortex and cerebellum respectively (Fig. 5e, Supplementary Table 8). They exhibited significant cortex- and cerebellum-specific expression patterns (Fig. 5f), suggesting that the region-specific differential editing may be explained by higher expression in their respective brain regions. It is likely that these region-specific genes have distinct functional roles in ASD.

We also examined 59 editing sites conserved across multiple phylogenetic taxa, likely serving as functionally paramount RNA editing sites in human[38]. Thirteen were identified as differentially edited in at least one brain region. Strikingly, they all exhibited hypoediting in ASD, 6 of which were recoding sites (Fig. 5g). Four of the recoding sites are located in glutamate receptors: *GRIA2* (R764G), *GRIA4* (R765G), *GRIK1* (Q621R) and *GRIK2*

(Y571C)[13]. Additionally, another recoding site was found in the *NOVA1* gene (Fig. 5g), which codes for a brain-specific splicing factor that reportedly may cause down-regulated splicing in ASD[39]. This recoding site (S363G) stabilizes protein half-life of NOVA1[39], suggesting that the down-regulated editing may be an upstream causal factor of down-regulated splicing in ASD[9]. Overall these findings strengthen the association between RNA editing and aberrant synaptic signaling in ASD.

Common and brain region-specific mechanisms of RNA editing regulation in ASD

Next, we examined the prospective regulation of hypoediting in the other brain regions. The eigen-editing sites of the turquoise modules in the other two brain regions also displayed correlations with both *FMR1* and *FXR1* expression (Fig. 5c, although the correlation for *FXR1* in cerebellum was not statistically significant, p = 0.07), suggesting that regulation of RNA editing by FMRP and FXR1P may be a common mechanism for multiple afflicted brain regions in ASD.

Correlation of the expression levels of the *ADAR* (1, 2 and 3) and Fragile X-related genes with the 1st principal component (PC) of all differential editing sites (Supplementary Fig. 28) also recapitulated many of the turquoise module associations: *FMR1* significantly associated with the 1st PC in both frontal cortex and cerebellum, and *FXR1* negatively correlated in all 3 brain regions, corroborating their roles in positive and negative regulation of RNA editing respectively. Although we did not observe significant changes of the *ADAR* mRNAs between ASD and control groups in any brain region (Supplementary Fig. 28), *ADAR2* was significantly associated with the 1st PC of differential editing in temporal cortex (Supplementary Fig. 28) and validated by Western blot analysis showing a possible trend of downregulated ADAR2 protein in the temporal cortex of ASD (Fig. 5h,i). Lastly, *FXR2*, though not associated with the turquoise module in frontal cortex, showed significantly positive correlation with the turquoise module in temporal cortex (Fig. 5c) and with the PC of differential editing in cerebellum (Supplementary Fig. 28). Future studies are needed to examine the roles of FXR2 and ADAR2 in these brain regions.

Exacerbated severity of hypoediting patterns in dup15q patients

Duplication of chromosome 15q11.2-q13.1 (i.e., dup15q), accounting for 0.25–3% ASD diagnoses[40], clinically manifests with more severe motor impairments and intellectual disability than idiopathic ASD[40,41], along with greater magnitude and homogeneous dysregulation of gene expression and splicing[9]. We analyzed RNA editing in dup15q from frontal cortex (8 samples), temporal cortex (9 samples), and cerebellum (5 samples) against covariate matched controls (Supplementary Fig. 29). Dup15q patients exhibited more profound hypoediting (Fig. 6a) than idiopathic ASD (Fig. 1c, 5a). Correlation between differential editing levels and the Intelligent Quotient (IQ) scores of the idiopathic ASD individuals (Supplementary Fig. 30) was also very high, though not significant because only a handful of ASD subjects had IQ information: temporal cortex (R^2=0.64), cerebellum (R^2=0.36), and most prominently in frontal cortex (R^2=0.80), the region considered most strongly associated with cognitive function[42]. These results suggest that editing dysregulation could be related to the severity of cognitive deficits.

The landscape of editing in dup15q recapitulated the trends in idiopathic ASD. Differential editing levels in dup15q significantly correlated with those in idiopathic ASD (Fig. 6b), and were enriched in the turquoise modules observed in the idiopathic subjects (Fig. 6c), and showed greater concordance and magnitude of hypoediting (Fig. 6d). We found hypoediting at nearly all the testable (Methods) 59 conserved sites, including replicated differential editing at the glutamate receptors *GRIA2* (R764G), *GRIA4* (R765G), *GRIK1* (Q621R), *GRIK2* (Y571C), and *NOVA1* (Fig. 6e). Overall, these results not only strongly validate the hypoediting landscape identified across the 3 brain regions of ASD but also reveal an exacerbated hypoediting bias in a subset of ASD patients with severe clinical phenotypes.

Discussion

Here we performed the first global investigation of RNA editing in ASD and uncovered a common trend of hypoediting in ASD patients across different brain regions and different patient cohorts. Furthermore, we showed correlation between the hypoediting and *FMR1* and *FXR1* genes, which we validated as direct regulators of multiple and diverse sites in human. Consistent with these roles, we demonstrated convergent RNA editing patterns between ASD and Fragile X syndrome, revealing a shared molecular deficit in these closely related neurodevelopmental disorders.

As the cause of the Fragile X syndrome and as a syndromic ASD, FMRP has been subject to a myriad of ASD studies: 1) genes with rare de novo mutations[17], common variation[43], and copy number variants[44] in ASD are enriched in FMRP target genes[30]. 2) Multiple transcriptome analyses identified significant correlation between FMRP expression and ASD-associated gene modules[8,18]. 3) Many similar cognitive and behavioral symptoms manifest in both ASD and Fragile X syndrome[36]. 4) The protein level of FMRP has been shown to be downregulated in ASD patients[45]. The plethora of related literature supports the involvement of FMRP in the pathogenesis of ASD and highlights the need to elucidate its potential molecular mechanisms. Our study addresses this question by showing that RNA editing may be strongly associated with the molecular pathology via which FMRP contributes to the molecular abnormalities observed in ASD.

Our data supports a model where FMRP directly mediates the interaction between ADAR and the RNA substrates to promote editing, which advances previous studies of FMRP and RNA editing in Mouse, Drosophila, and Zebrafish[29]. The involvement of FXR1P in RNA editing regulation was unknown, and intriguingly, we observed that FXR1P, likely through a similar model, represses editing. Additionally, FMRP and FXR1P showed distinct features among the validated regulatory targets, where FXR1P acted on promiscuous sites and FMRP on site-selective editing sites. Nevertheless, the two proteins also shared common validated target sites, suggesting they could have some synergistic regulation of RNA editing, as they do in other biological processes relevant to neurodevelopmental disorders, such as neurogenesis[29,46].

Our study revealed substantial similarities and highly reproducible patterns in global editing changes in ASD across the three brain regions we profiled, indicating it may affect molecular pathways in general neurological function. Nevertheless, our data also allude to some region-specific editing regulation, such as a downregulation trend of ADAR2 protein in the temporal cortex, but not in the frontal cortex or cerebellum. Expression levels of the gene *FXR2*, a homolog of *FMR1*, demonstrated strong correlation with RNA editing levels, which is again a temporal cortex-specific observation (Fig. 5c). Future studies aimed at studying region-specific RNA editing will further elucidate these and other region-specific regulatory mechanisms.

Individuals with ASD frequently score lower in IQ testing than neurotypicals[47]. Our analyses, although based on a small data set, showed a high correlation between differential editing and IQ scores in all 3 brain regions. Additionally, dup15q patients, generally known to manifest more severe motor impairments and intellectual disability than idiopathic ASD, showed nearly unidirectional and greater severity of hypoediting than idiopathic patients in all 3 brain regions. These findings support an association between intellectual disability and RNA editing in ASD, which awaits confirmation in subsequent cohorts.

RNA editing alterations occurred in genes of critical neurological relevance (Supplementary Fig. 11), including contactins (*CNTNAP2*, *CNTNAP4*), neurexins (*NRXN1*, *NRXN3*), ankyrins (*ANK2*), and neuronal splicing factors (*NOVA1* and *RBFOX1*), which all harbor genetic mutations associated with ASD[25]. Although causality here is indeterminable, the occurrence of aberrant RNA hypoediting in known ASD risk genes suggests these changes contribute to disease risk. They certainly contribute to the disorder's molecular pathology. Additionally, the differential editing sites significantly overlapped with developmentally regulated editing sites, suggesting that hypoediting may disrupt editing dependent functions during cortical developmental and further accentuates the potentiating role of early-onset molecular pathologies in ASD. Furthermore, some differential editing sites showed correlated editing levels with expression levels of their host genes, which may indicate a functional relationship. Together, this current work indicates that it will be important to further explore the role of RNA editing in ASD pathophysiology, so as to determine whether these changes are causal, or reflect homeostatic or dyshomeostatic responses.

Methods

RNA-Seq data sets of ASD and control brain samples

We obtained RNA-Seq data sets of three brain regions of ASD and control subjects from our previous study[9]. For idiopathic ASD, we used all data sets except (1) samples from subjects < 7 years old (which showed outlying expression patterns compared to all other samples), and (2) samples containing a 15q duplication (dup15q), an established genetic cause of syndromic ASD[48]. Note that the dup15q samples were analyzed separately as described below. We confirmed that ASD diagnosis was not confounded by age, batch, and other biological and technical variables (Supplementary Fig. 2). The final sample set consisted of an approximately equal number of controls and ASD samples totaling 62 samples in frontal cortex, 57 samples in temporal cortex, and 60 samples in cerebellum (Supplementary Table 1).

Dup15q dataset

A total of 5, 8, and 9 RNA-Seq datasets of dup15q patients were obtained from cerebellum, frontal cortex, and temporal cortex respectively (Supplementary Table 1). In addition, 11, 14, and 13 controls were chosen respectively from the above idiopathic dataset to balance covariates (Supplementary Fig. 29), except batch and brain bank, as there were nominally significant (although not passing Bonferroni correction) confounding effects between batch, brain bank, and dup15q diagnosis for this subset of data[9].

Frontal cortex replication dataset

For replication of idiopathic results, we downloaded previously published RNA-Seq data that were obtained from frontal cortex of 63 ASD and control subjects[27]. After balancing confounding variables, 22 ASD and 23 control datasets remained, none of which overlapped subjects from our original dataset.

RNA-Seq data sets of Fragile X patients and carriers/controls

Postmortem frontal cortex samples of Fragile X patients and Carriers were obtained from the University of Maryland Brain and Tissue Bank and the University of California at Davis FXTAS Brain Repository (Supplementary Fig. 25). Total RNA was extracted using TRIzol (Thermo Fisher Scientific, 15596018). RNA-Seq libraries were prepared using NEBNext Poly (A) mRNA magnetic isolation module (NEB, E7490) followed by NEBNext Ultra Directional RNA library prep kit for Illumina according to manufacturer's instruction. RNA-Seq data were collected on an Illumina HiSeq 2000 sequencer.

RNA-Seq read mapping and RNA editing identification

RNA-Seq reads were mapped using RASER[49], an aligner optimized for detecting RNA editing sites, using parameters $m = 0.05$ and $b = 0.03$. Uniquely mapped read pairs were retained for further analysis. Unmapped reads were extracted and processed to identify "hyperediting" sites. A recent study showed that previous RNA editing identification methods failed to detect editing sites in hyperedited regions due to existence of a high number of mismatches in the reads[20]. Our implementation of the hyperediting pipeline closely followed a strategy used by a previous study[20]. In brief, all adenosines in unmapped reads were converted into guanosines. These reads were aligned to a modified hg19 genome where adenosines were also substituted by guanosines. Uniquely mapped read pairs were obtained from this alignment step, and previously converted adenosines were reinstituted. We then combined these hyperedited reads with the originally uniquely mapped reads to identify RNA editing sites.

The procedures described in our previous studies were used to identify RNA editing sites[19,50]. First, RNA editing sites were identified as mismatches between the reads and the human reference genome. A log-likelihood test was carried out to determine whether an RNA editing site is likely resulted from a sequence error[19]. A number of posterior filters were then applied to remove RNA editing sites that were likely caused by technical artifacts in sequencing or read mapping[50].

Due to limited sequencing depth and the inherent nature of random sampling in RNA-Seq, some editing sites are observable in only a small number of subjects within a population cohort. Editing sites with low apparent prevalence lack sufficient sample size to enable a comparison between ASD and control groups. Therefore, we applied the following filters to retain a subset of editing sites: (1) in each sample, an editing site was required to have at least 5 total reads among which at least 2 reads were edited; (2) the editing site should satisfy filter (1) in at least 5 samples. We applied these filters to editing sites called within each brain region separately.

Identification of differential RNA editing sites

We define differential RNA editing sites as those: (1) that had significantly different average editing levels between ASD and controls, or (2) that were observed at significantly different population frequencies. A challenge with statistical testing for differential editing levels is that editing level estimation has a larger uncertainty at lower read coverage. More accurate calculations could be obtained by setting a high threshold for read coverage. However, this remedy leads to fewer samples or reduced power per editing site. We developed a strategy that attempts to optimize the trade-off between read coverage requirement and sensitivity in detecting differential editing.

Specifically, the following procedures were implemented for each editing site e_i. (1) we first identified the highest total read coverage requirement for e_i at which there were at least 5 samples in both control and ASD groups. The following read coverage was considered: 20, 15 and 10, in the order of high to low. (2) If a read coverage requirement C was reached in (1), we calculated the average editing level of e_i among the ASD and control samples (M_{ai}, M_{ci}), respectively, that satisfied C. (3) We then considered samples where e_i did not have at least C reads, but satisfied a lower read coverage cutoff (15, 10, or 5). These samples were included if their inclusion did not alter M_{ai} and M_{ci} by more than 0.03. (4) We carried out Wilcoxon rank-sum test between editing levels of the above samples in ASD and control groups. (5) If a read coverage requirement C was not reached in (1), then we tested all samples where e_i had ≥ 5 read coverage so long as there were at least 10 ASD and 10 control samples. Differential editing sites were defined as those with a p value < 0.05 and an effect size $> 5\%$, in lieu that an editing change of approximately this magnitude was sufficient to cause dendritic deficits in mice[51].

Another type of differential editing was defined as editing sites that have different prevalence between ASD and controls. For each editing site, a Fisher's Exact test was carried out to compare the numbers of ASD and control samples with nonzero editing levels, with the background being the numbers of ASD and control samples with zero editing level. The minimum read coverage requirement per site was obtained using the same adaptive procedure as described above for the first type of differential editing sites. Differential editing sites were defined as those with $p < 0.05$ and an effect size $> 5\%$. Differential editing sites identified via the above two methods overlapped significantly (Supplementary Fig. 7).

Differential editing sites detection in the replication ASD dataset[27] was performed similarly. However, because only 4952 editing sites had sufficient coverage, we eschewed effect size cutoffs and considered sites differential in which $p < 0.05$.

The dup15q subset sample size was too small to leverage the adaptive coverage model. Instead, we only tested editing sites where ≥ 5 dup15q and ≥ 5 control samples had ≥ 5 read coverage (defined as tesSupplementary Table ites). Differential editing sites had $p < 0.05$ and effect size > 5% from either Wilcoxon rank-sum test or Fisher's Exact test.

Computational comparison of methods and parameters for differential editing identification

Another *de facto* method for conducting differential testing in postmortem brain studies is to leverage a multilinear regression model to correct for potential technical confounders. We compared the results of our methods against those of a multilinear regression model including diagnosis, sex, age, and RIN as independent factors against RNA editing level. The set of differential editing sites strongly and significantly overlapped across all brain regions and available sample sizes (Supplementary Fig. 8, odds ratio 7–139), suggesting that *a priori* balancing of ASD and control groups was sufficient to obviate technical conflation. An additional issue with multilinear regression is a propensity for spuriously introducing noise at editing sites with smaller training sizes. Indeed, we found that the differential editing sets at smaller sample sizes (0–10 and 10–20) had more disparate calls between the two methods than the larger samples sizes (20–60).

We also tested whether the particular choice of parameters chosen for M_{ai} and M_{ci} significantly altered the differential editing values. We performed differential editing analysis with varying values of M_{ai} and M_{ci}, and juxtaposed the differential editing values with the originally called values (M_{ai} and $M_{ci} = 0.03$) (Supplementary Fig. 9). The correlation remained nearly at 1, which shows that the differential editing values are robust to the choice of M_{ai} and M_{ci}.

Identification of genes enriched with differential editing

This analysis aims to identify genes that are enriched with differential editing sites. One might consider the top differentially edited genes as those with the largest number of differential editing sites. However, as expected, there exists a positive correlation between gene length and the number of differential editing sites (Supplementary Fig. 11). Therefore, we used a linear model to construct a regression between these two variables. We defined genes as enriched with differential editing if they had more differential editing sites than expected (beyond 95% confidence interval of the expected mean).

Differential editing sites associated with gene expression

To examine the association between differential editing and gene expression, we screened for significant correlations between editing level of each differential editing site and the FPKM value of its host gene. Specifically, the correlation coefficient between editing level and FPKM had to pass nominal significance ($P < 0.05$) within a multilinear regression: FPKM ~ age, sex, batch, RIN, brain bank, seqStatPC1, seqStatPC2, editing level. seqStatPC1 and seqStatPC2 are the first and second principal components encompassing 99% of variance of technical variables as described in our previous work[9].

Enrichment of editing sites in developmentally distinct editing clusters

Editing sites identified in 33 postmortem frontal cortex samples spanning the human lifespan (fetal, infant, child, teen, middle, and old age) were obtained from a previous study[26]. The original study classified editing sites into 3 developmental trajectories (constantly lowly edited sites, perpetually highly edited sites, and developmentally increasing sites). We recapitulated the 3 developmental trajectories on editing sites residing in *all* genomic regions using similar clustering criteria as in the original study. Briefly, editing sites with a median coverage < 20 reads across all samples were discarded. Then, we performed one-way ANOVA on each editing site across the six age groups. We considered editing sites passing FDR < 0.05 as developmentally increasing sites. Amongst the remaining sites, those with median editing level > 0.5 were categorized as perpetually highly edited sites; those with median < 0.5, as constantly lowly edited sites. Enrichment of editing sites in ASD within these 3 developmental clusters was performed using Fisher's Exact test.

Annotation of editing sites and heatmap generation

Editing sites were annotated using ANNOVAR[52]. Heatmaps throughout this study were generated using circlize[53].

Principal components analysis (PCA)

PCA was conducted on differential editing sites in order to examine associations between PCs and potential confounding covariates. The R function prcomp was used for this purpose. Missing values in the editing level matrix were imputed using the missMDA package[54]. The PCs were then correlated against technical and biological covariates such as age and gender (Supplementary Fig. 10). The first PC was predominantly associated with ASD diagnosis, and was thus used as the PC for differential editing.

Weighted gene co-expression network analysis (WGCNA)

The WGCNA package[28] in R was used to find modules of correlated editing sites. In multi-sample analysis, it is typical that some editing sites have no available values (missing data) in certain samples that lack read coverage at those sites. To preclude inaccurate calculations due to samples with too much missing data, we used the following requirements for editing sites to be included in WGCNA: (1) with ≥5 reads in ≥ 90% of samples and (2) with nonzero editing levels in ≥ 10% samples. In addition, to detect variation in the data, we further required that the included editing sites had a standard deviation ≥ 0.1 in their editing levels across samples. A soft threshold power of 10 was used to fit scale-free topology. To avoid obtaining modules driven by outlier samples, we followed our previous bootstrapping strategy[9,11]. One hundred bootstraps of the data set were carried out to compute the topological overlap matrix of each resampled network. Co-editing modules were obtained using the consensus topological overlap matrix of the 100 bootstraps.

WGCNA offers a dynamic tree-cutting algorithm, which enables identification of modules at various dendrogram heights and allows delineation of nested modules[55]. However, upon examination of the WGCNA dendrogram (Fig. 2d), we observed only one pronounced module of editing sites. Furthermore, most modules, identified through dynamic tree cutting, were generally unstable, highly dependent on tree cutting parameters. Therefore, we used the traditional constant height tree cutting, provided by WGCNA as cutreeStaticColor, with cutHeight set to 0.9965, which produced the single turquoise module. This is the largest module that is most likely biologically relevant and technically robust. In addition, this module is conserved across brain regions (Fig. 5c).

Association of modules with ASD diagnosis and RNA binding proteins

To test the association of the turquoise module with diagnosis, we first defined eigen-editing sites as the first principal component of the module, according to WGCNA recommendation[56]. A linear regression model was constructed between the eigen-editing sites and diagnosis, in addition to biological and technical covariates including RIN, age, gender, sequencing batch, PMI, brain bank, 5' to 3' RNA bias, AT dropout rate, GC dropout rate, mapped bases in intergenic regions, uniquely mapped reads. The linear model was fit with backwards selection, and the module was deemed as associated with ASD diagnosis if $p \leq 0.05$ for the coefficient of this variable.

We tested if a module was enriched with differential editing sites using Fisher's Exact test. In addition, we tested the association between modules and potential regulatory genes by examining the correlation between the eigen-editing sites and mRNA expression levels of the genes. It should be noted that the mRNA expression levels were corrected values following removal of variability contributed by technical covariates[9].

eCLIP-Seq experiment and data analysis

The eCLIP experimental procedure is detailed in our previous studies[32,57]. The antibodies used for this experiment are: FMRP antibody (MBL, RN016P) and FXR1 antibody (Bethyl Laboratories, A303–892A). Flash-frozen brain tissue was cryo-ground in pestles pre-chilled with liquid nitrogen, spread out in standard tissue culture plates pre-chilled to −80°C, and UV crosslinked twice at 254 nM (400 mJ/cm^2). 50 mg of crosslinked tissue was then used for each eCLIP experiment, performed as previously described[32,57]. As controls, we sampled 2% of the pre-immunoprecipitation (post-lysis and fragmentation) sample and prepared libraries identically to the FMRP or FXR1P eCLIP (including the membrane size selection step). These libraries served as "size-matched input (SMInput)" to minimize non-specific background signal in the identical size range on the membrane as well as any inherent biases in ligations, RT-PCR, gel migration and transfer steps.

eCLIP-Seq data were analyzed using the CLIPper software[32] that generated a list of predicted binding peaks of the corresponding protein. In each replicate, peaks were further filtered to retain those whose abundance was at least 2 fold of that in the SMInput sample.

To examine the FMPR or FXR1P binding relative to RNA editing sites, we compared the distances from eCLIP peaks to turquoise editing sites compared to gene-matched random adenosines. Only editing sites residing in genes containing at least 1 eCLIP peak were considered. The closest distance between an editing site or random adenosine and eCLIP peaks were calculated. A total of 10,000 sets of controls were generated using this procedure. To determine a P value, we first plotted the cumulative distribution of the distances between editing sites or controls and the eCLIP sites. The area under the curve (AUC) of this distribution was calculated for the set of editing sites and each set of controls.

The AUC calculation was constricted to the distance interval [0,100,000 kb]. AUC values of the 10,000 sets of controls were modeled by a Gaussian distribution, which was used to calculate a P value for the AUC of the set of editing sites. Density plots were generated using the geom_density function in the ggplot2 package in R. To avoid overplotting, we randomly selected and plotted ten of the control sets for visualization. Note that the observed linear distance between protein-RNA binding and the regulated target sites may be larger than the actual proximity of the protein and its targets, due to limited sensitivity of CLIP or existence of secondary or tertiary RNA structures.

To identify the motifs enriched in eCLIP peaks, we used two alternative methods: HOMER[58], and DREME[59]. We ran DREME with all eCLIP peaks of each protein using default parameters, which creates control sequences through dinucleotide shuffling. HOMER was run with the findMotifsGenome.pl program (-p 4 -rna -S 10 -len 5,6,7,8,9). Background controls were defined as randomly chosen sequences in the same type of genic region as the true peaks. The control sequences have one-to-one match in length with the actual peaks. Three sets of random controls were constructed. Homopolymer or dinucleotide repeats were discarded. We required the final consensus motif to be the most enriched motif identified by HOMER that was also one of the most enriched motifs resulting from DREME.

RNA editing analysis of Fragile X samples

The RNA-Seq data derived from Fragile X patients and carriers were analyzed similarly as those of the ASD cohorts. Fisher's Exact test was used to identify differentially edited sites using pooled patient and carrier data sets ($p \le 0.05$ and effect size > 5%).

Gene ontology enrichment analysis

Gene ontology (GO) terms were downloaded from Ensembl[60]. For each query gene, a random control gene was chosen to match gene expression level and gene length (±10% relative to that of query gene). GO terms present in the sets of query genes and control genes were collected respectively. A total of 10,000 sets of control genes were obtained. For each GO term, a Gaussian distribution was fit to the number of control genes containing this GO term. The enrichment p value of the GO term among the query genes was calculated using this distribution.

Validation of RNA editing levels

RNA Extraction.—Brain tissues were homogenized in RNA TRIzol reagent (Thermo Fisher Scientific, 15596018). 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 12000g for 15 min, and the top layer was carefully extracted. Equal volume of 200-proof ethanol was added to the top chloroform layer and mixed thoroughly. RNA was further purified using Direct-zol™ RNA MiniPrep Plus kit (Zymo Research, R2072) following the manufacture's protocol.

cDNA synthesis and PCR.—cDNA synthesis was carried out using random hexamers, 1 μg total RNA, and the SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific, 18090050) as described in the manufacturer's protocol. Next, 2uL cDNA (corresponding to 1/10th of the original RNA) was used as template for PCR reactions using the DreamTaq PCR Master Mix (2X) (Thermo Fisher Scientific, k1082). PCR was performed on an Eppendorf thermal cycler using the following thermal cycle conditions for all candidate sites (5 min, 95°C for hot start followed by 30 cycles of 15 s, 95°C; 15 s, 55°C and 1min/kb, 72°C).

Topo Cloning and Clonal Sequencing.—PCR products were run on 1% agarose gel and visualized under UV light. The correct size band was cut and digested by Zymoclean™ Gel DNA Recovery Kit (Zymo Research, D4002) following the manufacturer's protocol. 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. Each plate was divided into 4 quadrants and 6 white clones were randomly selected from each quadrant (total of 24 clones per patient sample per editing site). Each clone was inoculated overnight in LB/Kanamycin. Plasmid was extracted using Plasmid DNA Miniprep Kits (Thermo Fisher Scientific, K210011). Miniprep samples were sequenced using Genewiz Sanger sequencing. The number of the clones presenting G peak at the editing site of interest was counted to determine the estimated editing ratio.

Co-immunoprecipitation

HeLa cells were maintained with DMEM supplemented with 10% FBS and 100 U ml-1 penicillin/ streptomycin at 37 °C and 5% CO2. Ten million HeLa cells were collected and lysed in 1 ml non-denaturing lysis buffer at pH 8.0, containing 20 mM Tris-HCl, 137 mM NaCl, 1% NP-40, and 2 mM EDTA supplemented with complete protease inhibitor cocktail. Extracted proteins were incubated overnight with ADAR1 antibody (Santa Cruz, sc-271854) or FMRP antibody (Millipore, MAB2160) at 4 °C; precipitation of the immune complexes was performed with Dynabeads Protein G (Thermo Fisher Scientific, 1003D) for 4 h at 4 °C, according to the manufacturer's instructions. For experiments involving Flag-ADAR2, the supernatant derived from Flag-tagged hADAR2 overexpressing cells was incubated for 3 h at 4 °C with Flag M2 antibody (Sigma, F1804). After immunoprecipitation, the beads were washed three times with the lysis buffer at 4 °C, and eluted from the Dynabeads using elute buffer (0.2 M glycine, at pH 2.8). Twenty microliters were loaded onto the gel and the samples were processed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and analyzed by Western blot. The following antibodies were used for the Western blots: ADAR1 antibody (Santa Cruz, sc-73408), Flag antibody (sc-807), FMRP antibodies (Millipore, MAB2160 and Abcam, ab17722), FXR1P antibody (Bethyl Laboratories, A303–892A), and FXR2 antibody (Sigma-Aldrich, F1554). The HRP-linked secondary antibodies were used and the blots were visualized with the ECL kit (GE, RPN2232).

Subcellular fractionation

Cells were fractionated following a previously published protocol with some modifications[61]. Briefly, monolayers of cells in 10-cm plates were washed twice with ice-cold PBS, followed by gentle scraping of cells. Cells were resuspended with the ice-cold

HLB+N buffer (10 mM Tris-HCl, at pH 7.5, 10 mM NaCl, 2.5 mM MgCl2 and 0.5% NP-40) on ice for 5 min. Lysates were layered over a chilled 10% sucrose cushion made in the ice-cold HLB+NS buffer (10 mM Tris-HCl, at pH 7.5, 10 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40 and 10% sucrose) and centrifuged for 5 min at 4 °C at 420g. After centrifugation, the supernatant was collected and served as the cytoplasmic fraction. The nuclear pellet was then treated with the ice-cold nuclei lysis buffer (10 mM HEPES, at pH 7.6, 300 mM NaCl, 7.5 mM MgCl2, 0.2 mM EDTA, 1 mM DTT, 1 M Urea, and 1% NP-40) after washing. Fractionation efficiency was validated by Western blot using antibody specific to the marker for each fraction: β-tubulin (Sigma, T8328) for the cytoplasmic fraction and rabbit polyclonal U1–70k (Santa Cruz, sc-390899) for the nucleoplasmic fraction.

Construction of minigenes and site-directed mutagenesis

Partial 3’ UTRs (EEF2K and TEAD1) and intronic (CNTNAP4, NLGN1, and TENM2) regions were restriction digested and inserted between the SacII/XhoI sites in the pEGFP-C1 vector. Overlapping oligonucleotide primers containing the desired mutations were used to amplify mutation-containing fragments from the wild-type minigene plasmid, using Q5 High-Fidelity DNA polymerase (New England Biolabs, M0491L). All resulting amplification products were confirmed by sequencing.

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

HeLa cells were grown on 6-well plates under standard conditions at 37 °C with 5% CO2. Cells were grown to 70% confluence, and transfection was performed using Lipofectamine 3000 (Thermo Fisher Scientific, L3000015) with 100 ng of minigene plasmid. For editing validation of endogenous substrate, two neuroblastoma cell lines, SK-N-BE(2) and KELLY, were grown on 6-well plates without transfection of a minigene. Cells were harvested after 24 h. Total RNA was extracted using TRIzol reagent (Thermo Fisher Scientific, 15596018), followed by treatment with 1 U of DNase I (Zymo Research, E1011-A). RNA was further purified using Direct-zol RNA MiniPrep kit following the manufacture’s instruction (Zymo Research, R2072). Reverse transcription (RT) was performed on 1 μg total RNA for 1 h at 42 °C using random hexamer primer and SuperScript IV (Thermo Fisher Scientific, 18090050). The cDNA products derived from the expressed minigenes were detected by PCR using the pEGFP-C1-specific forward primer and a gene-specific reverse primer. On the other hand, cDNA products for the endogenous substrate were amplified with gene-specific primer set. Amplification was performed for 30 cycles, consisting of 30 s at 95 °C, 30 s at 55 °C, and 2 minutes at 72 °C. The products from RT-PCR were resolved on 0.8% agarose gels. The appropriate PCR product was excised and the DNA was extracted, purified, and analyzed by Sanger sequencing. A-to-I editing levels were calculated as relative peak heights (that is, ratio between the G peak height and the combined height of A and G peaks, height G / height A + height G).

Production of lentivirus and cell transduction for protein knockdown

pLKO1 non-target control-shRNA (SHC016), FMR1-targeting shRNA (TRCN0000059758) or FXR1-targeting shRNA (TRCN0000159153) constructs 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[62]. Briefly, viruses were collected from conditioned media after 48 h co-transfection. Lentivirus-containing media was mixed with the same volume of DMEM media that contain polybrene (8 μg/ml), which was used to infect HeLa, SK-N-BE(2), and KELLY cells. After 24 h, cells were incubated with puromycin (2 μg/ml for HeLa and 1 μg/ml for SK-N-BE(2) and KELLY) for 3–7 days. Knockdown efficiency was evaluated by Western blot. Cells were lysed in RIPA containing complete protease inhibitor cocktail. Cell lysates were then resolved through 8% SDS-PAGE and probed by ADAR1 antibody (Santa Cruz, sc-271854), ADAR2 antibody (Santa Cruz, sc-73409), FMRP antibody (Millipore, MAB2160), FXR1P antibody (Bethyl Laboratories, A303–892A), and FXR2 antibody (Sigma-Aldrich, F1554).

Western Blot in ASD and Fragile-X brain samples

Brain tissues were homogenized in RIPA lysis and extraction buffer containing protease inhibitor (Thermo Scientific, 88866). Mixture was then incubated on ice for 30 minutes, sonicated, and spun down. Crude protein concentration was obtained using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, 23225). Equal amount of protein was separated using 8% SDS–PAGE and then transferred onto nitrocellulose membrane. The membrane was blocked with 5% non-fat milk (Genesee Scientific, 20–241) and 0.1% Tween 20 in tris-buffered saline. The blot was incubated in primary antibody solution against the protein of interest with 5% non-fat milk and 0.1% Tween 20 in TBS overnight at 4°C on shaker. Antibodies used in this experiment include ADAR1 antibody (Santa Cruz, sc-271854), ADAR2 antibody (Santa Cruz, sc-73409), ADAR3 antibody (Santa Cruz, sc-73410), FMRP antibody (Millipore, MAB2160). Secondary antibody containing goat anti-mouse IgG-HRP (sc-2005, Santa Cruz Biotechnology) or goat anti-rabbit IgG-HRP (sc-2004, Santa Cruz Biotechnology) was used to label the corresponding primary antibody. The blot was developed using Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences, RNP2232) and imaged with the Syngene PXi immunoblot imaging system. Beta Actin was used as a loading control. Western blot images were analyzed using ImageJ. All uncropped images are included in Supplementary Fig. 31.

RNA immunoprecipitation (RIP)–PCR

RIP was performed according to previously published protocols with some modifications[63]. Cells were harvested on the second day of minigene transfection in RIP buffer (25 mM Tris-HCl, at pH 7.4, 150 mM KCl, 5 mM EDTA, 0.5% NP-40 and 0.5 mM DTT supplemented with complete protease inhibitor cocktail and 100 U ml-1 RNase OUT), sonicated (10 s three times with 1 min intervals) and centrifuged at 13,000 rpm for 10 min at 4 °C. Supernatant was treated with 100 U RNase-free DNase I (Zymo Research, E1011-A) at 37 °C for 30 min and then centrifuged again at 13,000 rpm for 10 min at 4 °C. For immunoprecipitation, lysates were incubated with FXR1P antibody (Santa Cruz, sc-374148) or anti-mouse IgG (Santa Cruz, sc-2025) as a negative control overnight at 4 °C. The Dynabeads were washed three times with the RIP buffer and bound RNA was isolated using TRIzol (Thermo Fisher Scientific, 15596018), according to the manufacturer's instructions. Eluted RNA was reverse-transcribed using SuperScript IV (Thermo Fisher Scientific, 18090050) with random hexamer primers. PCR was carried out for 30 cycles, consisting of

30 s at 95 °C, 30 s at 55 °C, and 30 s at 72 °C. PCR products were analyzed by agarose gel electrophoresis.

Immunofluorescence

HeLa cells were seeded on Millicell EZ Slide 8-well glass (Millipore, PEZGS0816) and incubated overnight in DMEM with 10% FBS to obtain 60% monolayer cell confluency. Each chamber was carefully rinsed with ice-cold PBS. Cells were fixed in 4% paraformaldehyde at room temperature for 10 min and washed with ice-cold 0.1% PBS-T three times for total of 15 min. Cells were permeabilized with either 0.1% Tween-20 or Triton X-100 in PBS for 5 min. Block solution containing 5% normal donkey serum and 1% BSA in 0.3% PBS-T was applied for 1 h at room temperature on shaker. Cells were incubated in primary antibody solution of mouse anti-ADAR1 (1: 100; sc-271854, Santa Cruz Biotechnology) and rabbit anti-FMR1 (1: 100; ab17722, Abcam) in 0.3% PBS-T containing 1% NDS and 1% BSA for overnight at 4°C. Cells were washed three times with ice-cold 0.1% PBS-T for 5 min. Cells were then incubated in a secondary cocktail containing Highly Cross-Adsorbed AlexaFluor® 488-conjugated Donkey anti-Mouse IgG (1: 200; A-21202, Thermo Fisher Scientific), and AlexaFluor® 488-conjugated Donkey anti-Rabbit IgG (1: 200; ab150074, Abcam) in 0.3% PBS-T containing 1% NDS and 1% BSA. Chamber was disassembled to expose the slide. Vectashield Anti-fade mounting medium containing 4’,6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI) stain was applied to the slide and covered with a coverslip. Cells were examined and imaged at 63x oil-immersion objective using Zeiss LSM 780 confocal microscope with ZEN 2011 (Black edition) software and post-processed with ImageJ. All images were taken under identical setting and conditions.

Statistics

Differential editing sites were obtained using a two-tailed Wilcoxon signed-rank test under an adapting scheme (see previous section). Ascertaining bias for hypoediting was performed using a Chi-square test under the null hypothesis of equal numbers of up- and down-regulated editing sites. Significance of gene set and editing set overlaps were determined using a two-tailed Fisher’s exact test. Significance of minigene reporter assays were summarized using one-way ANOVA and a Student’s t-test against proper controls, where data distributions were assumed to be normal, but this was not formally tested. Data generated in this study was not randomized according to experimental conditions or stimulus presentations, and data collection and analyses were not performed blind to the conditions of the experiments. For statistics of more specific analyses, see the appropriate sections in Methods and Figure legends (also refer to the online “Life Sciences Reporting Summary”).

Sample size selection

No statistical methods were used to pre-determine sample sizes, but our samples sizes are similar to those reported in previous publications.[8,9]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Postmortem brain samples of used in this study were obtained from the University of Maryland Brain and Tissue Bank, which is a component of the NIH NeuroBioBank. We are grateful to the patients and families who participate in the tissue donation programs. This work was funded by grants from the National Institute of Health to X.X. (HG009417, HG006264), G.W.Y. (HG004659, HG009417, HG007005 and MH107367), S.T. (T32HG002536), E.L.V.N. (HG009530), V.M.C. (MH094681), R.J.H. (HD 036071). S.T is supported by the UCLA Eureka Scholarship. E.L.V.N. is a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2172–13). G.A.P. is supported by the National Science Foundation Graduate Research Fellowship. G.R. is supported by NIH fellowship 1F32MH114620.

References

1. Association, A. P. Diagnostic and statistical manual of mental disorders (4th *ed.,* text rev). 4th edn, (2000).
2. Rojas DC The role of glutamate and its receptors in autism and the use of glutamate receptor antagonists in treatment. Journal of neural transmission (Vienna, Austria : 1996) 121, 891–905, doi: 10.1007/s00702-014-1216-0 (2014).
3. Guo YP & Commons KG Serotonin neuron abnormalities in the BTBR mouse model of autism. Autism research : official journal of the International Society for Autism Research 10, 66–77, doi: 10.1002/aur.1665 (2017). [PubMed: 27478061]
4. Ha S, Sohn IJ, Kim N, Sim HJ & Cheon KA Characteristics of Brains in Autism Spectrum Disorder: Structure, Function and Connectivity across the Lifespan. Experimental neurobiology 24, 273–284, doi:10.5607/en.2015.24.4.273 (2015). [PubMed: 26713076]
5. Nelson SB & Valakh V Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. Neuron 87, 684–698, doi:10.1016/j.neuron.2015.07.033 (2015). [PubMed: 26291155]
6. Marchetto MC et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. Mol Psychiatry, doi:10.1038/mp.2016.95 (2016).
7. de la Torre-Ubieta L, Won H, Stein JL & Geschwind DH Advancing the understanding of autism disease mechanisms through genetics. Nature medicine 22, 345–361 (2016).
8. Gupta S et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat Commun 5, 5748, doi:10.1038/ncomms6748 (2014). [PubMed: 25494366]
9. Parikshak NN et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature 540, 423–427, doi:10.1038/nature20612 (2016). [PubMed: 27919067]
10. Irimia M et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell 159, 1511–1523, doi:10.1016/j.cell.2014.11.035 (2014). [PubMed: 25525873]
11. Wu YE, Parikshak NN, Belgard TG & Geschwind DH Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. Nat Neurosci, doi:10.1038/nn.4373 (2016).
12. Nishikura K A-to-I editing of coding and non-coding RNAs by ADARs. Nature reviews. Molecular cell biology 17, 83–96, doi:10.1038/nrm.2015.4 (2016). [PubMed: 26648264]
13. Behm M & Ohman M RNA Editing: A Contributor to Neuronal Dynamics in the Mammalian Brain. Trends in genetics : TIG 32, 165–175, doi:10.1016/j.tig.2015.12.005 (2016). [PubMed: 26803450]
14. Slotkin W & Nishikura K Adenosine-to-inosine RNA editing and human disease. Genome medicine 5, 105, doi:10.1186/gm508 (2013). [PubMed: 24289319]
15. Khermesh K et al. Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. RNA (New York, N.Y.) 22, 290–302, doi:10.1261/rna.054627.115 (2016).
16. Eran A et al. Comparative RNA editing in autistic and neurotypical cerebella. Mol Psychiatry 18, 1041–1048, doi:10.1038/mp.2012.118 (2013). [PubMed: 22869036]
17. Iossifov I et al. De novo gene disruptions in children on the autistic spectrum. Neuron 74, 285–299, doi:10.1016/j.neuron.2012.04.009 (2012). [PubMed: 22542183]

18. Parikshak NN et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155, 1008–1021, doi:10.1016/j.cell.2013.10.031 (2013). [PubMed: 24267887]
19. Bahn JH et al. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome research 22, 142–150, doi:10.1101/gr.124107.111 (2012). [PubMed: 21960545]
20. Porath HT, Carmi S & Levanon EY A genome-wide map of hyper-edited RNA reveals numerous new sites. Nat Commun 5, 4726, doi:10.1038/ncomms5726 (2014). [PubMed: 25158696]
21. Picardi E, D'Erchia AM, Lo Giudice C. & Pesole G REDIportal: a comprehensive database of A-to-I RNA editing events in humans. Nucleic acids research 45, D750–d757, doi:10.1093/nar/gkw767 (2017). [PubMed: 27587585]
22. Levanon EY et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nature biotechnology 22, 1001–1005, doi:10.1038/nbt996 (2004).
23. Oakes E, Anderson A, Cohen-Gadol A & Hundley HA Adenosine Deaminase That Acts on RNA 3 (ADAR3) Binding to Glutamate Receptor Subunit B Pre-mRNA Inhibits RNA Editing in Glioblastoma. The Journal of biological chemistry 292, 4326–4335, doi:10.1074/jbc.M117.779868 (2017). [PubMed: 28167531]
24. Rakic P Evolution of the neocortex: a perspective from developmental biology. Nature reviews. Neuroscience 10, 724–735, doi:10.1038/nrn2719 (2009). [PubMed: 19763105]
25. Abrahams BS et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Molecular autism 4, 36, doi:10.1186/2040-2392-4-36 (2013). [PubMed: 24090431]
26. Hwang T et al. Dynamic regulation of RNA editing in human brain development and disease. Nat Neurosci 19, 1093–1099, doi:10.1038/nn.4337 (2016). [PubMed: 27348216]
27. Liu X et al. Disruption of an Evolutionarily Novel Synaptic Expression Pattern in Autism. PLoS biology 14, e1002558, doi:10.1371/journal.pbio.1002558 (2016). [PubMed: 27685936]
28. Langfelder P & Horvath S WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 9, 559, doi:10.1186/1471-2105-9-559 (2008). [PubMed: 19114008]
29. Davis JK & Broadie K Multifarious Functions of the Fragile X Mental Retardation Protein. Trends in genetics : TIG, doi:10.1016/j.tig.2017.07.008 (2017).
30. Darnell JC et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 146, 247–261, doi:10.1016/j.cell.2011.06.013 (2011). [PubMed: 21784246]
31. Zhang Y et al. The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. The EMBO journal 14, 5358–5366 (1995). [PubMed: 7489725]
32. Van Nostrand EL et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nature methods 13, 508–514, doi:10.1038/nmeth.3810 (2016). [PubMed: 27018577]
33. Ascano M, Jr. et al. FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature 492, 382–386, doi:10.1038/nature11737 (2012). [PubMed: 23235829]
34. Vasudevan S & Steitz JA AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. Cell 128, 1105–1118, doi:10.1016/j.cell.2007.01.038 (2007). [PubMed: 17382880]
35. Van Nostrand EL et al. A Large-Scale Binding and Functional Map of Human RNA Binding Proteins. bioRxiv, doi:10.1101/179648 (2017).
36. Hagerman R, Hoem G & Hagerman P Fragile X and autism: Intertwined at the molecular level leading to targeted treatments. Molecular autism 1, 12, doi:10.1186/2040-2392-1-12 (2010). [PubMed: 20858229]
37. Abbeduto L, McDuffie A & Thurman AJ The fragile X syndrome-autism comorbidity: what do we really know? Front Genet 5, 355, doi:10.3389/fgene.2014.00355 (2014). [PubMed: 25360144]
38. Pinto Y, Cohen HY & Levanon EY Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. Genome biology 15, R5, doi:10.1186/gb-2014-15-1-r5 (2014). [PubMed: 24393560]
39. Irimia M et al. Evolutionarily conserved A-to-I editing increases protein stability of the alternative splicing factor Nova1. RNA biology 9, 12–21, doi:10.4161/rna.9.1.18387 (2012). [PubMed: 22258141]

40. DiStefano C et al. Identification of a distinct developmental and behavioral profile in children with Dup15q syndrome. Journal of neurodevelopmental disorders 8, 19, doi:10.1186/s11689-016-9152-y (2016). [PubMed: 27158270]
41. Battaglia A et al. The inv dup(15) syndrome: a clinically recognizable syndrome with altered behavior, mental retardation, and epilepsy. Neurology 48, 1081–1086 (1997). [PubMed: 9109904]
42. Frith C & Dolan R The role of the prefrontal cortex in higher cognitive functions. Brain research. Cognitive brain research 5, 175–181 (1996). [PubMed: 9049084]
43. Jansen A et al. Gene-set analysis shows association between FMRP targets and autism spectrum disorder. European journal of human genetics : EJHG 25, 863–868, doi:10.1038/ejhg.2017.55 (2017). [PubMed: 28422133]
44. Pinto D et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. American journal of human genetics 94, 677–694, doi:10.1016/j.ajhg.2014.03.018 (2014). [PubMed: 24768552]
45. Fatemi SH & Folsom TD Dysregulation of fragile x mental retardation protein and metabotropic glutamate receptor 5 in superior frontal cortex of individuals with autism: a postmortem brain study. Molecular autism 2, 6, doi:10.1186/2040-2392-2-6 (2011). [PubMed: 21548960]
46. Patzlaff NE, Nemec KM, Malone SG, Li Y & Zhao X Fragile X related protein 1 (FXR1P) regulates proliferation of adult neural stem cells. Human molecular genetics 26, 1340–1352, doi: 10.1093/hmg/ddx034 (2017). [PubMed: 28204491]
47. Charman T et al. IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). Psychological medicine 41, 619–627, doi:10.1017/s0033291710000991 (2011). [PubMed: 21272389]
48. Finucane BM et al. in GeneReviews(R) (eds Adam MP *et al.*) (University of Washington, Seattle University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle All rights reserved., 1993).
49. Ahn J & Xiao X RASER: reads aligner for SNPs and editing sites of RNA. Bioinformatics (Oxford, England) 31, 3906–3913, doi:10.1093/bioinformatics/btv505 (2015).
50. Lee JH, Ang JK & Xiao X Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. RNA (New York, N.Y.) 19, 725–732, doi: 10.1261/rna.037903.112 (2013).
51. Feldmeyer D et al. Neurological dysfunctions in mice expressing different levels of the Q/R site-unedited AMPAR subunit GluR-B. Nat Neurosci 2, 57–64, doi:10.1038/4561 (1999). [PubMed: 10195181]
52. Wang K, Li M & Hakonarson H ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research 38, e164, doi:10.1093/nar/gkq603 (2010). [PubMed: 20601685]
53. Gu Z, Gu L, Eils R, Schlesner M & Brors B circlize Implements and enhances circular visualization in R. Bioinformatics (Oxford, England) 30, 2811–2812, doi:10.1093/bioinformatics/btu393 (2014).
54. Josse J & Husson F missMDA: A Package for Handling Missing Values in Multivariate Data Analysis 2016 70, 31, doi:10.18637/jss.v070.i01 (2016).
55. Langfelder P, Zhang B & Horvath S Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics (Oxford, England) 24, 719–720, doi:10.1093/bioinformatics/btm563 (2008).
56. Langfelder P & Horvath S Eigengene networks for studying the relationships between co-expression modules. BMC systems biology 1, 54, doi:10.1186/1752-0509-1-54 (2007). [PubMed: 18031580]
57. Wheeler EC, Van Nostrand EL & Yeo GW Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. Wiley interdisciplinary reviews. RNA, doi:10.1002/wrna.1436 (2017).
58. Heinz S et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589, doi: 10.1016/j.molcel.2010.05.004 (2010). [PubMed: 20513432]

59. Bailey TL, Johnson J, Grant CE & Noble WS The MEME Suite. Nucleic acids research 43, W39–49, doi:10.1093/nar/gkv416 (2015). [PubMed: 25953851]

60. Aken BL et al. The Ensembl gene annotation system. Database : the journal of biological databases and curation 2016, doi:10.1093/database/baw093 (2016).

61. Nojima T, Gomes T, Carmo-Fonseca M & Proudfoot NJ Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. Nature protocols 11, 413–428, doi:10.1038/nprot.2016.012 (2016). [PubMed: 26844429]

62. Sundararaman B et al. Resources for the Comprehensive Discovery of Functional RNA Elements. Mol Cell 61, 903–913, doi:10.1016/j.molcel.2016.02.012 (2016). [PubMed: 26990993]

63. Bahn JH et al. Genomic analysis of ADAR1 binding and its involvement in multiple RNA processing pathways. Nat Commun 6, 6355, doi:10.1038/ncomms7355 (2015). [PubMed: 25751603]

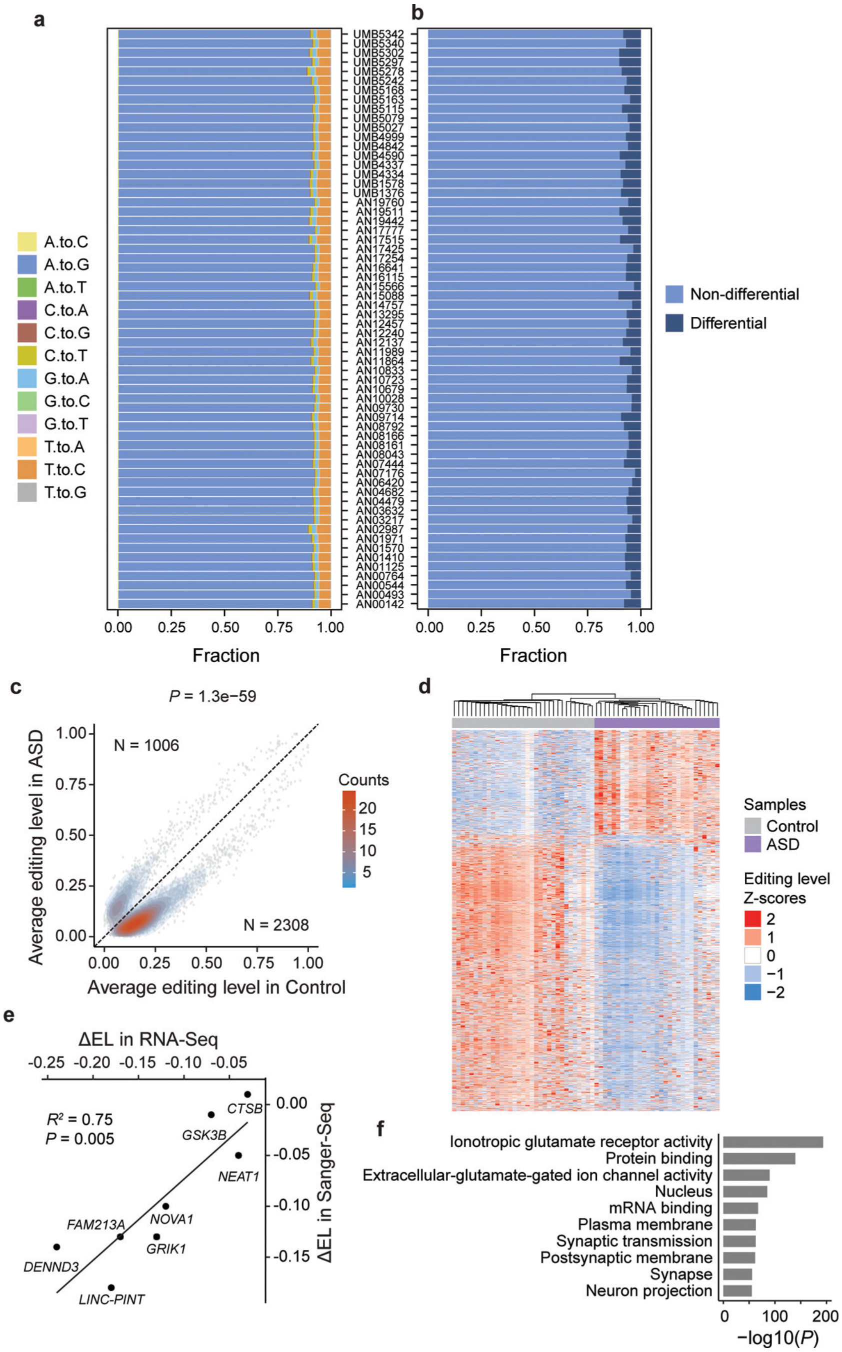

Figure 1. Transcriptome-wide differential editing in the frontal cortex of ASD.

a, Fraction of all types of RNA-DNA differences (RDDs) identified in the RNA-Seq data of each subject. **b**, Fraction of differential and non-differential editing sites for each subject. **c**, Average editing levels of differential editing sites in ASD and controls. Numbers (N) of editing sites that were up- or down-regulated in ASD are shown, which were compared via Chi-squared test (P value shown above plot). **d**, Differential editing sites segregate ASD and control samples. Normalized editing levels (z-scores) were used in hierarchical clustering. Each row corresponds to one editing site. Each column represents a sample. **e**, Experimental validation of differential editing levels using Sanger sequencing. The frontal cortex samples used in this experiment are shown in Supplementary Table 1. ΔEL: change in editing level (ASD - control), n=8 editing sites. **f**, GO enrichment analysis of genes harboring differential editing sites (n=1,189 genes, p-values determined by one-sided Gaussian test, see methods).

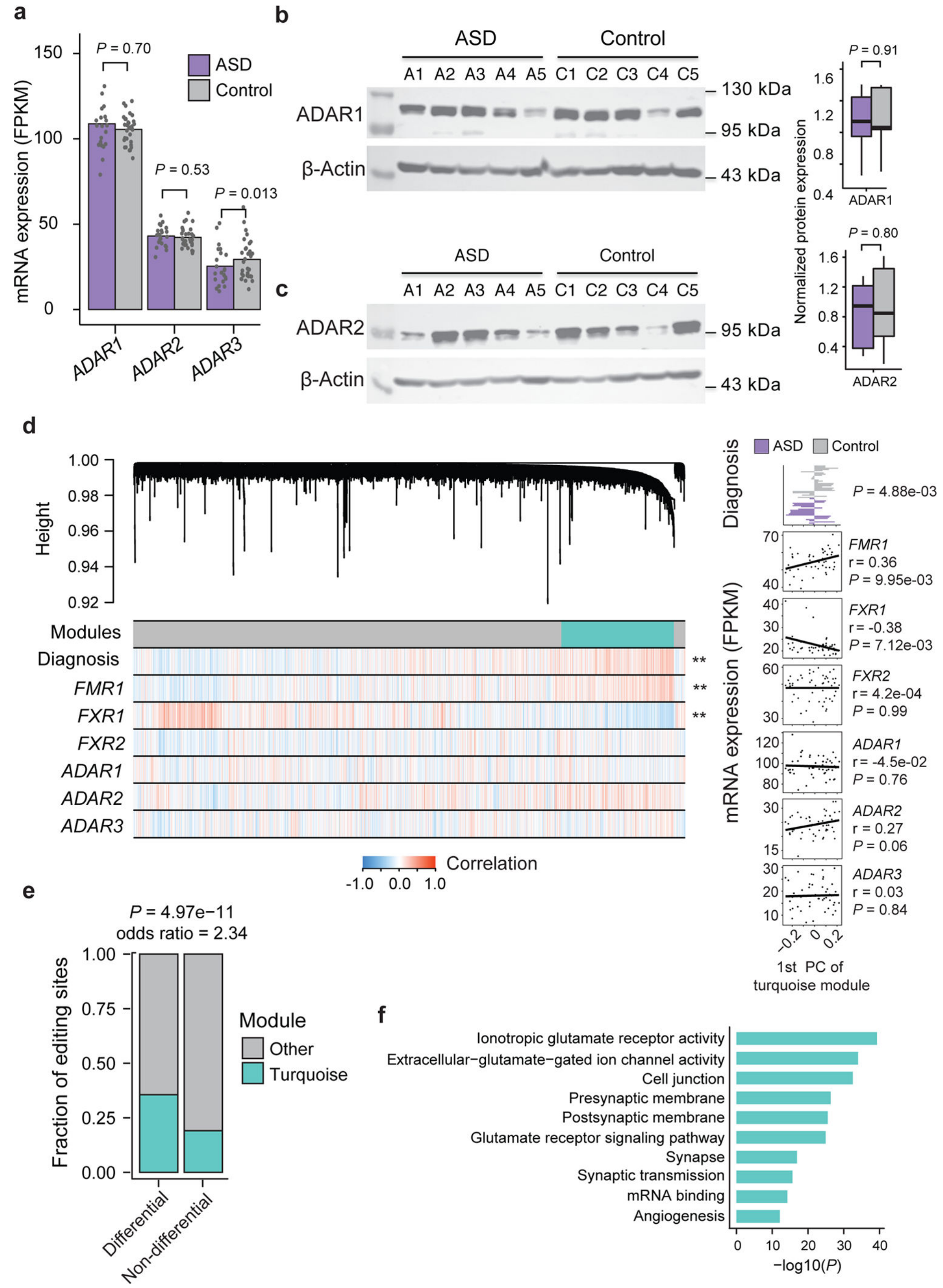

Figure 2. Global analysis reveals potential regulators of differential editing in the frontal cortex of ASD.

a, mRNA expression levels (FPKM) of ADAR1, ADAR2, and ADAR3 estimated from RNA-Seq data (n=62 samples). P values were calculated using a regression approach where covariates were accounted for[15]. Dots show individual sample FPKMs. **b**, Western blot of ADAR1 protein in ASD and control samples. (Note the images are cropped and uncropped images are in Supplementary Fig. 31, same for all Western blot images hereafter.) Protein level was normalized against that of β-actin. Samples used in this experiment are shown in

Supplementary Table 1 (chosen based on availability). A1-A5: ASD samples. C1-C5: control samples. P value was calculated via two-tailed Student's t-test. Boxplot definition: center=median, lower hinge=25th percentile, upper hinge=75th percentile, min and max extend to observations at most 1.5 * inter quartile range (IQR) . **c**, Similar as b, for ADAR2 protein. d, WGCNA analysis of RNA editing in frontal cortex (n=51 samples). Dendrogram of RNA editing sites is shown. The turquoise module is indicated by the turquoise color. Correlation of editing sites with diagnosis (ASD or control) and mRNA expression levels of a few genes is shown in the Heatmap. **: P < 0.01. Right panels: Bar graph and scatter plots represent association between diagnosis or mRNA expression levels and the first principal component (PC) of the turquoise module. P values of Pearson's correlation are shown. **e**, Overlap between the turquoise sites and differential editing sites. P value was calculated via Fisher's Exact test (n=4061 editing sites, two-tailed). **f**, GO enrichment analysis of genes harboring the turquoise sites (n=846 genes, one-tailed Gaussian test, see methods).

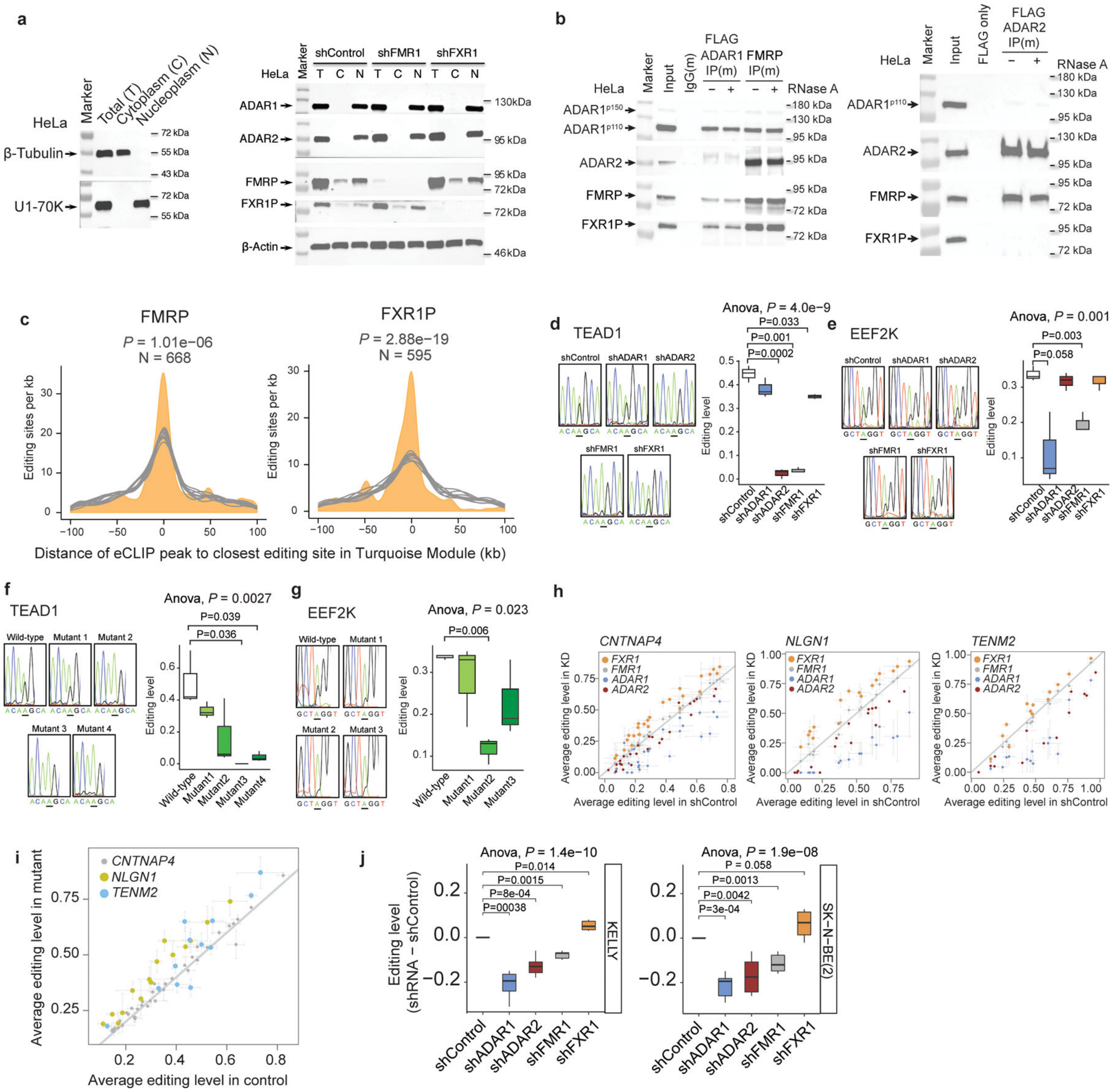

Figure 3. FMRP and FXR1P regulate RNA editing.

a, Western blot of ADAR1, ADAR2, FMRP and FXR1P proteins in the nuclear and cytoplasmic fractions of HeLa cells. Cell fractionation was confirmed by Western blotting of β-tubulin and U1–70K as marker proteins. Control cells and cells with stable knockdown of FMR1 or FXR1 were used. Experiment was repeated twice with similar results. **b**, Co-IP experiments with or without RNase A in HeLa cells between ADAR1, ADAR2, FMRP and FXR1P. Endogenous proteins were targeted except for ADAR2 (where a FLAG-tagged ADAR2 was expressed). Experiment was repeated 3 times with similar results. **c**, Shortest distance between FMRP or FXR1P eCLIP peaks and turquoise editing sites resulted from the WGCNA analysis (orange). Ten sets of random control sites (gray) are depicted for comparison (see Methods). Number of editing sites (N) is shown (see Methods for P value calculation). **d**, Experimental testing of an RNA editing site in the *TEAD1* gene for its dependence on ADAR1, ADAR2, FMRP or FXR1P. Control HeLa cells or cells with stable knockdown of one of these proteins were used to express a minigene that contains the editing site. Editing levels were measured by Sanger sequencing. Example sequencing traces are shown for each sample with the targeted editing site underlined. Boxplots include three biological replicates. Overall P value (shown above plot) was calculated by one-way ANOVA. Individual comparison P values were calculated by one-tailed Student's t-test. **e**, Similar as **d**, for an editing site in the *EEF2K* gene. f. similar as d, but displaying editing levels of the TEAD1 editing site in minigenes with the wild-type sequence or different versions of mutants introduced to predicted FMRP binding motifs (see Supplementary Fig. 20). Wild-type HeLa cells were used to express these minigenes. **g**, Similar as **f**, for the editing site in the EEF2K gene (see Supplementary Fig. 20). h, RNA editing levels in control HeLa cells and cells with stable knockdown of ADAR1, ADAR2, FMR1 or FXR1. Hyper-editing sites in three genes were tested. Error bars represent standard errors of three biological replicates. **i**, Editing levels in control HeLa cells in the same hyper-edited genes as in h, but with mutations introduced at predicted FXR1 binding motifs (see Supplementary Fig. 23c-e). Error bars are standard errors of three biological replicates. **j**, Editing changes at six differential editing sites in ASD induced by control shRNA (shControl) or shRNA knockdown of ADAR1, ADAR2, FMR1, and FXR1 in two neuroblastoma cell lines, KELLY and SK-N-BE(2) (see Supplementary Fig. 24). Boxplots show editing changes against shRNA control over the six editing sites (n=6 editing sites). P-values calculated using two-tailed t-test. Boxplot definitions for d-g, j: center=median, lower hinge=25th percentile, upper hinge=75th percentile, min and max extend to observations at most 1.5 * IQR.

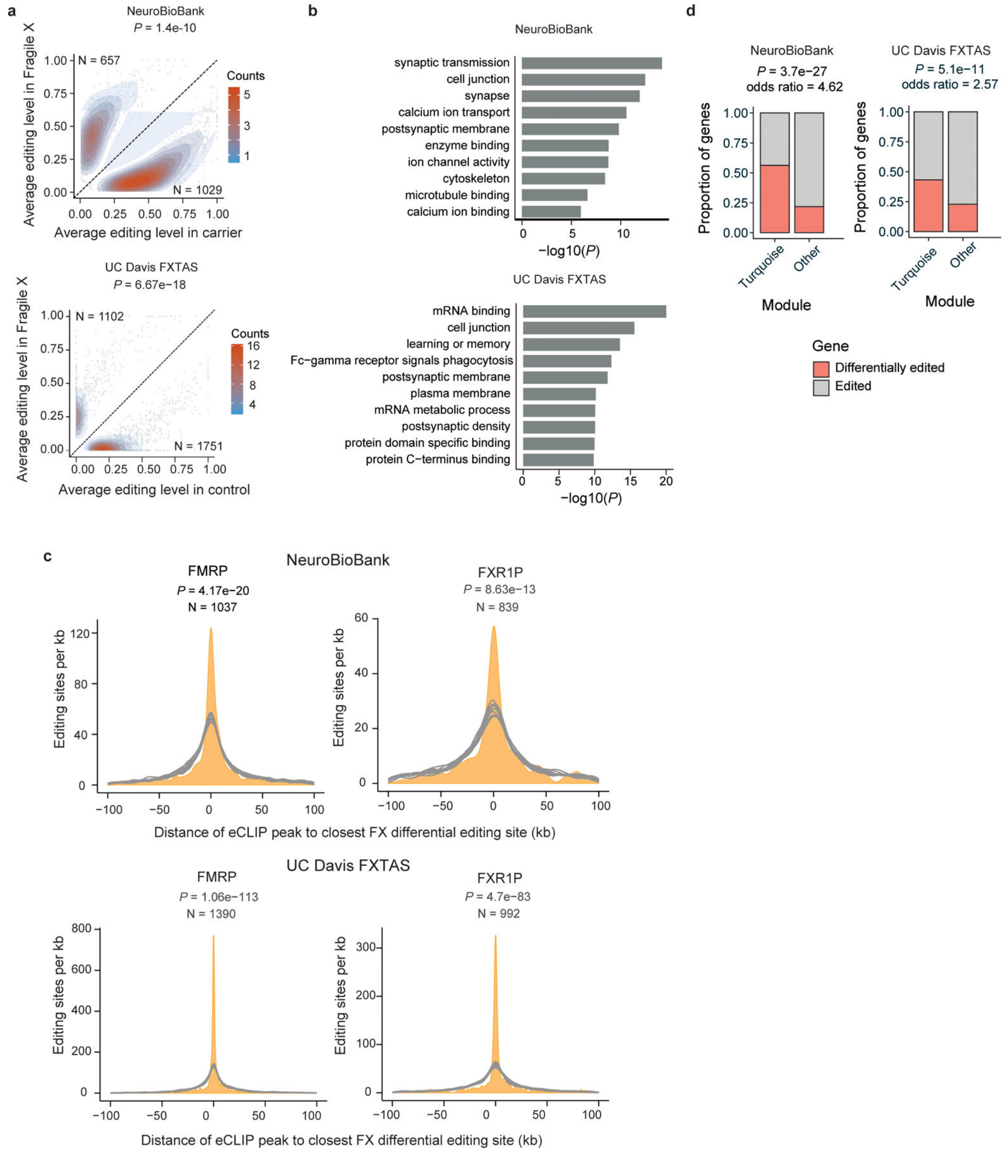

Figure 4. Transcriptome-wide differential editing in the frontal cortex of Fragile X patients and controls.

a, Average editing levels of differential editing sites in Fragile X patients compared to carriers (left; NeuroBiobank dataset) or controls (right; UC Davis FXTAS dataset). Numbers of editing sites (N) that were up- or down-regulated in the patients are shown, which were compared via Chi-squared test (P value shown above plot). **b**, Gene ontology enrichment of genes harboring differential editing sites (n=961 and 1914 genes for Neurobiobank and UC Davis FXTAS respectively; one-tailed Gaussian test, see methods). **c**, Similar as Fig. 3c, shortest distance between FMRP or FXR1P eCLIP peaks and differential editing sites in a (orange) (n=number of differential editing sites overlapping eCLIP genes; P-value from one-tailed Gaussian test, see methods). **d**, Overlap between the genes harboring WGCNA turquoise sites of ASD frontal cortex and those harboring differential editing sites in the Fragile X patients. P value was calculated via two-tailed Fisher's Exac test (N = 4051 and 7915 genes in Neurobiobank and UC Davis FXTAS respectively).

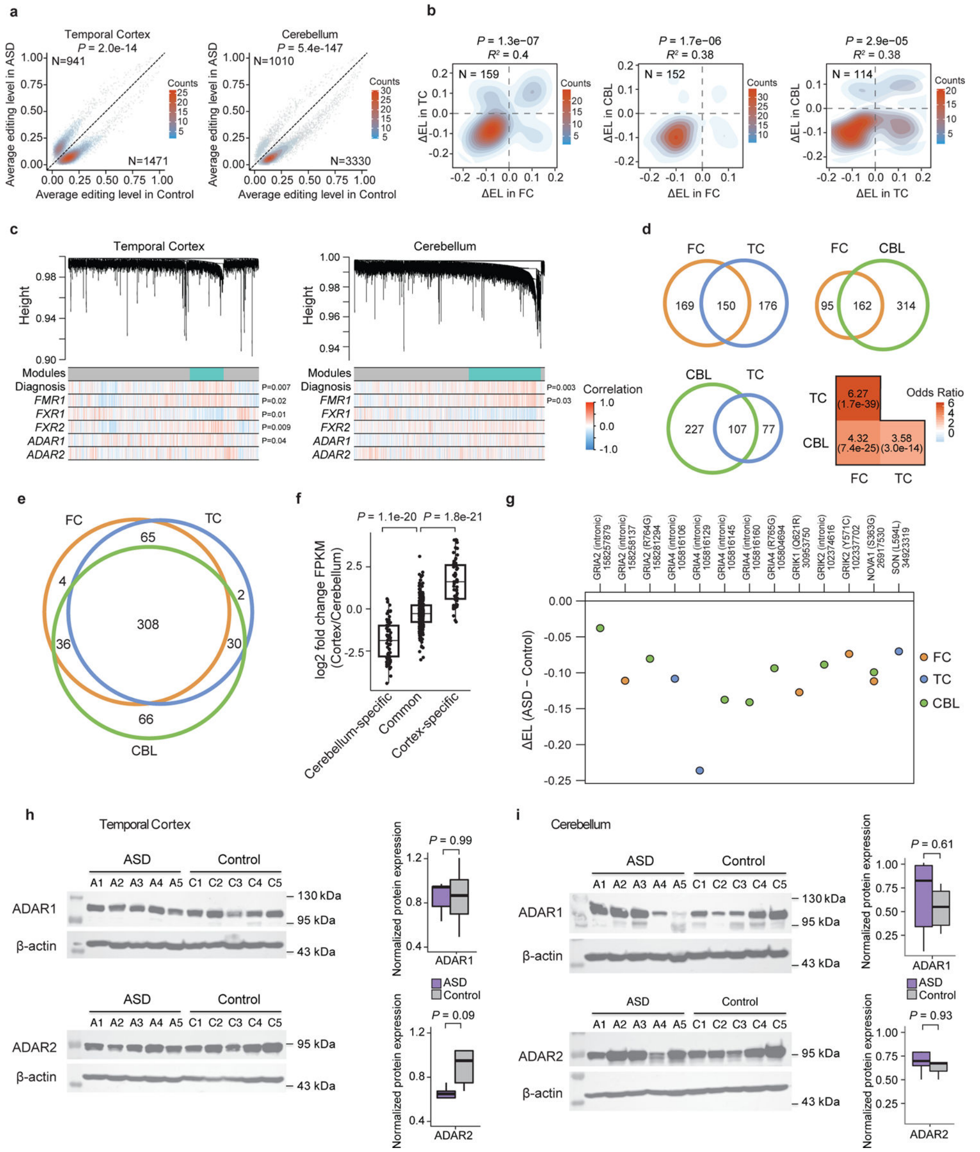

Figure 5. RNA editing dysregulation in different brain regions.

a, Average editing levels of differential editing sites in ASD and controls. Similar as Fig. 1c, but data for temporal cortex and cerebellum are shown respectively. N=number differential editing sites, P-value calculated by Chi-square test. **b**, Changes in editing levels (ΔEL) between ASD and control (ASD-control) of differential editing sites shared across brain regions (N=number of differential editing sites shared). Abbreviations (same below): TC: temporal cortex; FC: frontal cortex; CBL: cerebellum. Pearson's correlation P and R^2 values are shown. **c**, Similar as Fig. 2d, for WGCNA analysis of editomes in the temporal cortex and cerebellum regions. P-values calculated from linear regression (Methods). N = 46 and 47 samples in temporal cortex and cerebellum respectively. **d**, Overlap of editing sites in the turquoise modules of pairs of brain regions. Only editing sites with sufficient read coverage in both brain regions for WGCNA analysis are included. Odds ratios and P values (in parenthesis) for the overlaps are shown in the heatmap (two-tailed Fisher's Exact test). **e**, Overlap of genes harboring differential editing sites across brain regions. **f**, Relative FPKM values (log2 fold change) between cortex and cerebellum samples for genes that harbor differential editing sites only in cerebellum, only in cortex or in both types of regions. P values were calculated by two-tailed Student's t-test. N=66 cerebellum, 301 common, and 59 cortex specific genes respectively. **g**, Editing level difference (ΔEL, ASD-control) for a small number of literature-reported evolutionarily conserved RNA editing sites that showed differential editing between ASD and control groups in at least one brain region. **h**, Similar as Fig. 2b, Western blot of ADAR1 and ADAR2 proteins in temporal cortex samples. Samples used in this experiment are shown in Supplementary Table 1. N=10 samples. **i**, Similar as **h**, Western blot of ADAR1 and ADAR2 proteins in cerebellum samples. N = 10 samples. Boxplot definitions: center=median, lower hinge=25th percentile, upper hinge=75th percentile, min and max extend to observations at most 1.5 * IQR.

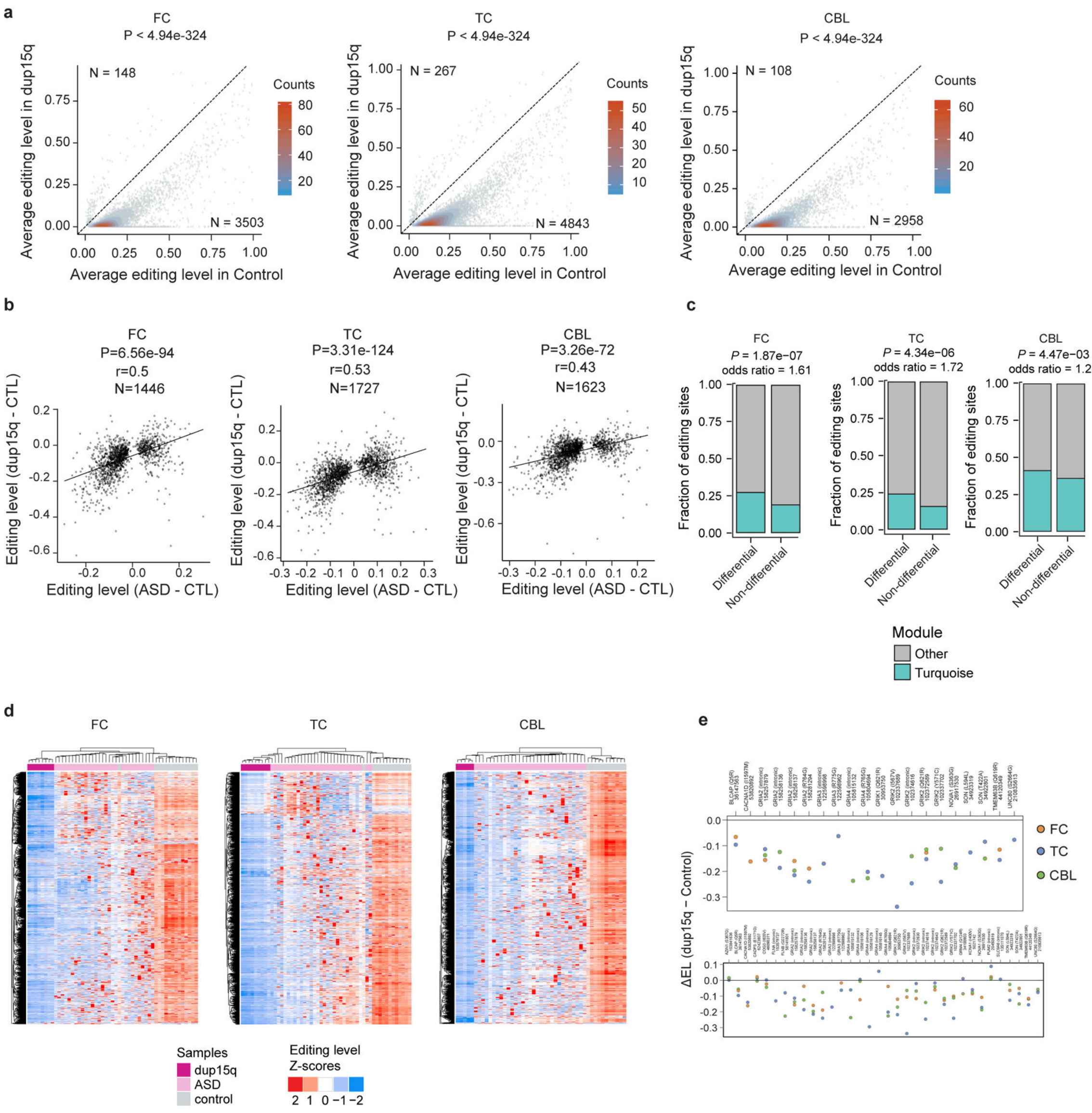

Figure 6. Hypo-editing in 3 brain regions of dup15q patients.

a, Contour scatterplot of differential editing sites in 3 brain regions of dup15q patients vs matched controls, similar as Fig. 1c. N=number of differential editing sites hypo and hyperedited. P-value calculated using Chi-square test. **b**, Comparison of editing level difference in idiopathic ASD and dup15q patients relative to their respective controls (CTL). Pearson correlation P values and correlation coefficient (r) are shown. N represents the number of editing sites that are differential in idiopathic ASD and testable (see Methods) in the dup15q patients. **c**, Overlap between differential editing sites in dup15q patients and the turquoise modules of the respective brain regions of idiopathic ASD. P values were calculated via two-tailed Fisher's Exact test, n=3411, 2224, and 3834 in FC, TC, and CBL respectively. **d**, Heatmaps (similar as Fig. 1d) of editing sites shown in b, including dup15q patients, matched controls, and the entire idiopathic ASD cohort. **e**, Editing level difference (ΔEL, dup15q-control) for a small number of literature-reported evolutionarily conserved

RNA editing sites. Top panel: differential sites in at least one brain region of dup15q patients. Bottom panel: all testable editing sites.