

Widespread RNA editing dysregulation in brains from autistic individuals

Stephen S. Tran^{1,2}, Hyun-Ik Jun², Jae Hoon Bahn², Adel Azghadi², Gokul Ramaswami³, Eric L. Van Nostrand^{4,5,6}, Thai B. Nguyen^{4,5,6}, Yun-Hua E. Hsiao⁷, Changhoon Lee³, Gabriel A. Pratt^{4,5,6,8}, Verónica Martínez-Cerdeño⁹, Randi J. Hagerman¹⁰, Gene W. Yeo^{4,5,6,8}, Daniel H. Geschwind^{11,12*} and Xinshu Xiao^{1,2,13,14*}

Transcriptomic analyses of postmortem brains have begun to elucidate 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 postmortem brains of people with ASD. 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 RNA-editing enzymes (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 data sets, including dup15q (genomic duplication of 15q11.2-13.1) cases associated with intellectual disability, highlight RNA-editing dysregulation in ASD and reveal new mechanisms underlying this disorder.

Autism spectrum disorder (ASD) is characterized by a developmental deficit in social communication accompanied by repetitive and restrictive interests¹, with a strong neuropathology implicating glutamatergic² and serotonergic³ circuits, aberrant structural development in multiple brain regions⁴, excitatory and inhibitory imbalance⁵, and abnormal synaptogenesis⁶. The genetic etiology of ASD remains incompletely understood and shows substantial heterogeneity⁷. Nevertheless, recent studies, leveraging the increasing availability of postmortem samples, have revealed shared patterns of transcriptome dysregulation affecting neuronal and glial coding and noncoding gene expression^{8,9}, neuronal splicing including microexons¹⁰, and microRNA targeting¹¹ in about two-thirds of people with ASD. These studies highlight downregulation of activity-dependent genes in neurons and upregulation of astrocyte and microglial genes as key points of convergence in ASD pathology.

Another major RNA processing mechanism is RNA editing, 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 genes¹². 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 has important roles in neurodevelopment and maintenance of normal neuronal function¹³. Indeed, a number of

A-to-I editing sites alone are imperative in modulating excitatory responses, permeability of ionic channels and other neuronal signaling functions¹³. Aberrant RNA editing has been reported in several neurological disorders, such as schizophrenia, bipolar disorder, amyotrophic lateral sclerosis¹⁴ and Alzheimer's disease¹⁵. In ASD, a previous study has analyzed a few known RNA editing sites in synaptic genes and reported altered editing patterns in a small cohort of ASD cerebella¹⁶. Yet it remains unaddressed whether global patterns of RNA editing contribute to the neuropathology of ASD, a question that requires larger subject cohorts and the study of 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 so far, spanning multiple brain regions. We identified a core set of downregulated RNA editing sites that are 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 proteins and modulate A-to-I editing. Mutations in FMRP lead to Fragile X syndrome, a disease with high comorbidity with ASD⁷. Indeed, we observed convergent dysregulated patterns of RNA editing in Fragile X and ASD subjects, consistent with findings that genes harboring ASD risk mutations are enriched in FMRP targets^{17,18}. Overall, we provide global insights

¹Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA, USA. ²Department of Integrative Biology and Physiology, UCLA, Los Angeles, CA, USA. ³Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA. ⁴Department of Cellular and Molecular Medicine, UCSD, La Jolla, CA, USA. ⁵Stem Cell Program, UCSD, La Jolla, CA, USA. ⁶Institute for Genomic Medicine, UCSD, La Jolla, CA, USA. ⁷Department of Bioengineering, UCLA, Los Angeles, CA, USA. ⁸Bioinformatics and Systems Biology Graduate Program, UCSD, La Jolla, CA, USA. ⁹Department of Pathology and Laboratory Medicine, UC Davis School of Medicine, Sacramento, CA, USA. ¹⁰The MIND Institute, Department of Pediatrics, UC Davis School of Medicine, Sacramento, CA, USA. ¹¹Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA. ¹²Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA. ¹³Molecular Biology Institute, UCLA, Los Angeles, CA, USA. ¹⁴Institute for Quantitative and Computational Biology, UCLA, Los Angeles, CA, USA.

*e-mail: dhg@mednet.ucla.edu; gxxiao@ucla.edu

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 postmortem subjects, we obtained ribosomal RNA-depleted total RNA sequencing (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 three brain regions and 20 subjects in common across two regions (Supplementary Fig. 1). These data sets were generated as part of our transcriptomic study of ASD brain⁹. Overall, the ASD and control groups did not have significant differences in variables that might confound RNA editing analysis (for example, age and gender) (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 data¹⁹, and implemented additional steps to capture editing sites located in ‘hyperedited’ regions, which were probably missed by regular methods²⁰ (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 we expected, the number of predicted RNA editing sites per sample was 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 and Supplementary Fig. 5). Notably, most (84%) of the A-to-I editing sites are listed in the REDIport database²¹ (Supplementary Fig. 5). The majority of RNA editing sites were located in Alu sequences²¹ (Supplementary Fig. 5) and in intronic regions²¹ (Supplementary Fig. 5), and the sequence context of A-to-G sites was consistent with the typical sequence signature known for ADAR substrates²² (Supplementary Fig. 5). Examination of correlation between ADAR expression levels and various partitions of editing sites (Alu, non-Alu-repetitive and nonrepetitive 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 literature²³, and together strongly support the validity of our predicted A-to-I editing sites.

The frontal and temporal cortex shared >70% of their sites, whereas 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, probably reflecting the substantial differences in cellular composition and physiology between these two regions²⁴.

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 $\geq 5\%$ or editing prevalence difference $\geq 5\%$, see Methods and Supplementary Table 2) that were robust to the choice of statistical models and parameters (Methods and Supplementary Fig. 7–9). For each individual, 2.6–10.5% of all editing sites were identified as differential (Fig. 1b). Notably, the differentially edited sites showed a bias of hypoediting in ASD samples (Fig. 1c); the number of downregulated RNA editing sites in ASD far outnumbered those that were upregulated ($P = 1.3 \times 10^{-59}$, chi-squared test, Fig. 1c).

Across potentially confounding biological and technical variables, diagnosis (that is, 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), indicating that differential editing was unlikely secondary to differential gene expression.

We used Sanger sequencing to confirm the observed editing differences of eight 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 (four ASD, four controls), selected based on sample availability (Supplementary Table 1b). The editing differences calculated from RNA-seq were 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 enrichment for categories including ionotropic glutamate receptor activity, glutamate gated ion channel activity and synaptic transmission (Fig. 1f). Consistently, genes (for example, KCNIP4, PCDH9, RBFOX1 and CNTNAP2) with the largest number of differential editing sites (both before and 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 genes²⁵ (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 development²⁶ (Supplementary Fig. 12). Together, these results indicate that RNA editing could contribute to aberrant synaptic formation in ASD.

Replication of hypoediting in an independent cohort of ASD frontal cortex. For replication, we analyzed an independent cohort of ASD subjects²⁷. After balancing technical covariates, we analyzed RNA-seq data from frontal cortex of 22 ASD and 23 controls (Supplementary Fig. 13 and Supplementary Table 1c). This data set had single-end reads from poly(A) 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 4,952 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 data set. Differential editing sites in the replication data set were likewise enriched in genes involved in synapse and cell junction (Supplementary Fig. 14), and the levels of differential editing were significantly correlated with those in our study (Supplementary Fig. 14). Replication of the editing landscape using data from a different cohort²⁷ collected by a different lab strongly supports the validity of our observed ASD editing profiles.

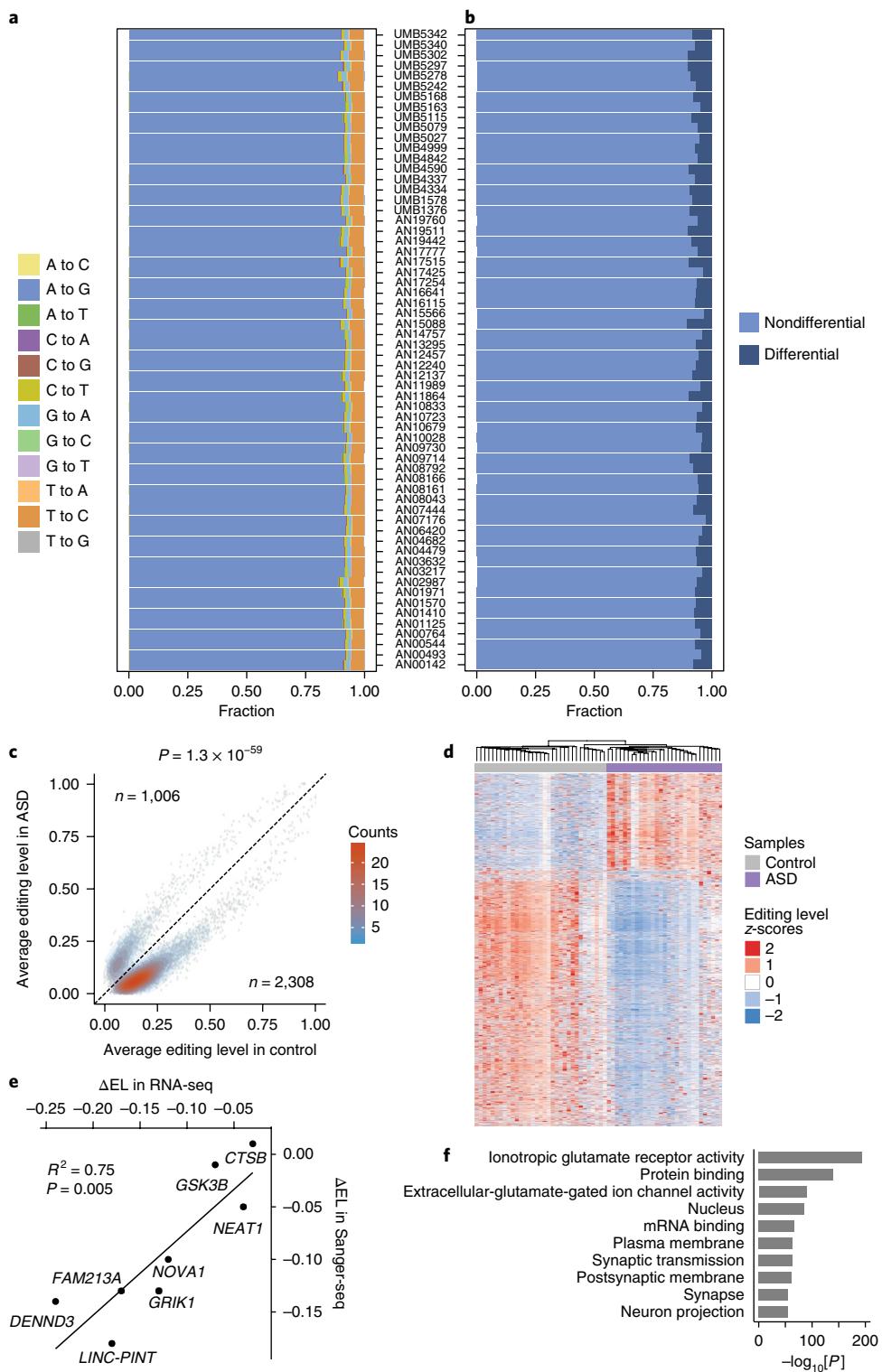


Fig. 1 | Transcriptome-wide differential editing in the frontal cortex of subjects with ASD. **a**, Fraction of all types of RNA-DNA differences (RDDs) identified in the RNA-seq data of each subject. **b**, Fraction of differential and nondifferential editing sites for each subject. **c**, Average editing levels of differential editing sites in ASD and controls. Numbers (n) of editing sites that were up- or downregulated 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**, Gene ontology enrichment analysis of genes harboring differential editing sites ($n=1,189$ genes, P values determined by one-sided Gaussian test; see Methods).

Global analysis of potential hypoediting regulators 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 *ADAR* (*ADAR1*) and *ADARB1* (*ADAR2*) expression in frontal cortex (Fig. 2a–c). Although *ADARB2* (*ADAR3*) protein was undetectable in the brain

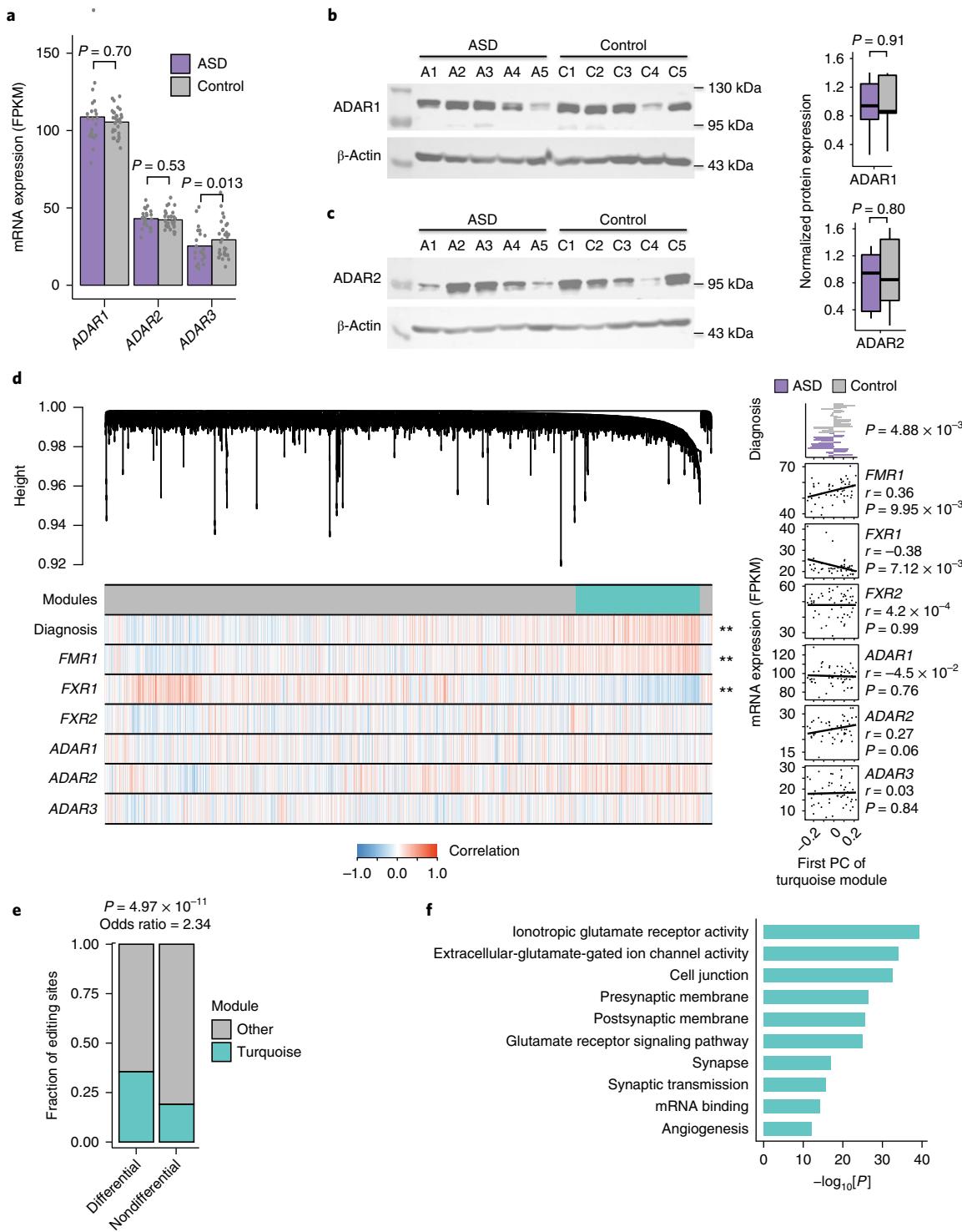


Fig. 2 | Global analysis reveals potential regulators of differential editing in the frontal cortex of ASD. **a**, mRNA expression levels of ADAR1, ADAR2 and ADAR3 estimated from RNA-seq data ($n=62$ samples). P values were calculated using a regression approach where covariates were accounted for¹⁵. Dots show individual sample fragments per kilobase per million mapped reads (FPKM) values. **b**, Western blot of ADAR1 protein in ASD and control samples. (Images are cropped, and uncropped images are in Supplementary Fig. 31, as for all western blot images hereafter.) Protein level was normalized against that of β -actin. Samples used in this experiment are 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, minimum and maximum extend to observations at most $1.5 \times$ interquartile range (IQR). **c**, Similar to **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 turquoise color. Correlation of editing sites with diagnosis (ASD or control) and mRNA expression levels of a few genes is shown in heatmap. ** $P < 0.01$. Right: 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 ($n=4,061$ editing sites, two-tailed). **f**, Gene ontology enrichment analysis of genes harboring the turquoise sites ($n=846$ genes, one-tailed Gaussian test; see Methods).

samples (Supplementary Fig. 15), its mRNA was slightly downregulated in ASD (Fig. 2a), which, as ADAR3 protein is an RNA editing inhibitor²³, cannot explain the observed hypoediting in ASD. The ADAR genes did not exhibit differential splicing in these samples, as has been determined previously⁹, and have no reported rare or common variants associated with ASD.

Given the absence of explanatory variation by ADAR genes, we hypothesized that 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 coexpression network analysis (WGCNA)²⁸ to search for highly correlated clusters of editing sites (that is, modules) (Methods).

Notably, we identified a module enriched in editing sites that had significant association with diagnosis (Fig. 2d and Supplementary Table 5), enriched with differential editing sites between ASD and controls in frontal cortex of this study (Fig. 2e), and enriched with differential editing sites from the replication cohort (Supplementary Fig. 14). Correlation between the module ‘eigengene’ (that is, 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 (that is enhancing) correlation with editing changes, whereas *FXR1* displayed negative (that is, 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 synapses²⁹, 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 a subcellular fractionation experiment followed by western blot and reciprocal co-immunoprecipitation 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 fraction¹², whereas FMRP and FXR1P were detected substantially in the cytoplasmic fraction³⁰ (Fig. 3a). Notably, FMRP and FXR1P were also highly detectable in nucleus; this was corroborated using immunofluorescence experiments (Supplementary Fig. 16). Subcellular distribution of ADAR proteins remained unchanged upon FMRP or FXR1P knockdown (Fig. 3a). Reciprocal co-immunoprecipitation experiments showed that FMRP interacted with both ADAR1 and ADAR2 in an RNA-independent manner (Fig. 3b), whereas 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 to dysregulated editing sites. Next, we captured the transcriptome-wide binding patterns of FMRP and FXR1P to RNA transcripts using enhanced ultraviolet cross-linking 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 and Supplementary Fig. 17).

We first confirmed the quality of our eCLIP experiments. eCLIP peaks identified in each replicate (Methods and 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' untranslated regions (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 previously reported FMRP binding motif³³, and CAUGC in FXR1P (Supplementary Fig. 17), which is consistent with a previous report that FXR1P tends to associate with (A+U)-rich elements³⁴.

Next, we examined the proximity of FMRP and FXR1P binding peaks relative to dysregulated editing sites in ASD frontal cortex. Notably, the FMRP and FXR1P eCLIP peaks were significantly enriched around editing sites in the turquoise module (Fig. 3c and Methods), a finding that was replicated in the FMRP eCLIP data generated from K562 cells by ENCODE³⁵ (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 and 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 probably site-specific editing sites, because 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 corroborate the positive association of the turquoise eigen-editing site with *FMR1* expression levels (Fig. 2d).

Next, we introduced mutations to the FMRP binding motifs in the minigenes 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 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 together²⁰. 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 and Supplementary Table 3), two of which (*CNTNAP4* and *NLGN1*) are ASD risk genes. Consistent with its known role in hyperedited RNA editing²⁰, *ADAR1* knockdown caused reduction in all the detectable editing sites (Fig. 3h), though *ADAR2* knockdown did too to a lesser degree. Notably, the hyperedited sites in these genes showed increased editing levels in *FXR1* (but not in *FMR1*) knockdown cells; this was again consistent with the WGCNA results that showed negative correlation between *FXR1* expression and RNA editing (Fig. 2d). RNA immunoprecipitation experiments supported the idea 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

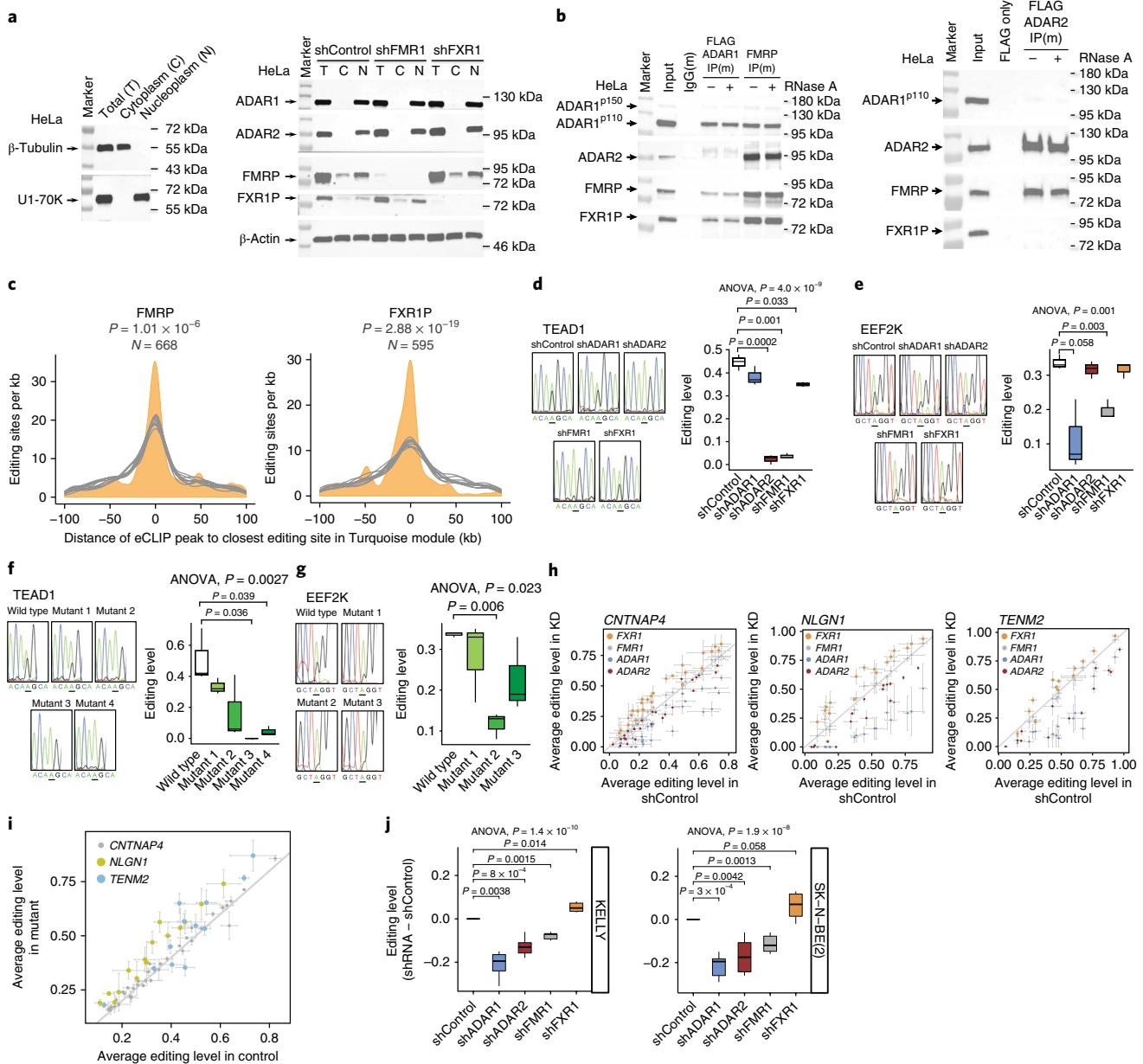


Fig. 3 | FMRP and FXR1P regulate RNA editing. **a**, Western blot of ADAR1, ADAR2, FMRP and FXR1P proteins in nuclear and cytoplasmic fractions of HeLa cells. Cell fractionation was confirmed by western blotting of β -tubulin and U1-70K as marker proteins. Control cells (shControl) and cells with stable knockdown of *FMR1* (shFMR1) or *FXR1* (shFXR1) were used. Experiment was repeated twice with similar results. **b**, Co-immunoprecipitation 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 three times with similar results. IP(m), immunoprecipitation with antibody derived from mouse. **c**, Shortest distance between FMRP or FXR1P eCLIP peaks and turquoise editing sites resulted from 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 to **d** but for an editing site in the *EEF2K* gene. **f**, Similar to **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 to **f** but 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 (KD) of ADAR1, ADAR2, *FMR1* or *FXR1*. Hyperediting 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 hyperedited genes as in **h**, but with mutations introduced at predicted FXR1P 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; minimum and maximum extend to observations at most $1.5 \times$ IQR.

minigenes (Fig. 3*i* and Supplementary Fig. 23). These results potentially indicate that FXR1P carries out direct inhibitory regulation of hyperediting sites through mediated interaction between ADAR1 and the RNA.

Concomitant regulation of RNA editing by FMRP and FXR1P. To further substantiate the above findings, we validated six more differential editing sites in two neuroblastoma cell lines (Supplementary Table 3). These candidate sites were chosen based on their proximity to FMRP or FXR1P eCLIP sites and their nominal correlation with the turquoise module or with *FMR1* or *FXR1* gene expression. As we expected, *ADAR1* and *ADAR2* short hairpin RNA (shRNA) knockdown reduced editing at all editing sites (Supplementary Fig. 24). Notably, *FMR1* shRNA knockdown caused significant reduction of editing of all sites, whereas *FXR1* knockdown caused significant augmentation of editing in 10 of the 12 sites (Fig. 3*j* and Supplementary Fig. 24). 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 validate that FMRP and FXR1P are important regulators of RNA editing.

Convergent RNA editing between ASD and Fragile X subjects. Loss of FMRP manifests in Fragile X syndrome, the most prevalent monogenic cause of ASD (1–2% of all ASD)^{7,36} in which ~50% of subjects have co-diagnoses or features of ASD³⁷. To investigate whether RNA editing contributes to shared molecular pathologies, we generated RNA-seq data from the frontal cortex of four subjects 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 analysis 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).

Notably, differential editing sites identified in the Fragile X data set (Supplementary Fig. 25 and Methods) showed the same trends as those from ASD: they demonstrated a predominant trend of hypoediting in Fragile X subjects and strong enrichment in genes related to synaptic transmission, cellular junctions and ionic transmission (Fig. 4*a,b* and Supplementary Table 7), and were also significantly enriched around FMRP and FXR1P eCLIP peaks (Fig. 4*c*). Moreover, a statistically significant overlap was observed between the differentially edited genes in Fragile X subjects and those in the turquoise module identified from ASD frontal cortex (Fig. 4*d*), the module that is correlated with *FMR1* expression (Fig. 2*d*). In addition, a significant overlap exists between the differential editing sites in Fragile X subjects 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 subject 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 in different brain regions of ASD subjects. Here we investigated whether other brain regions share similar editing patterns with the frontal cortex. In temporal cortex and cerebellum, we also observed global downregulation of RNA editing and enrichment in synapses, cellular junctions and ionic channels (Fig. 5*a* and Supplementary Fig. 27). Overall, differential editing sites shared between brain regions showed significant correlation

in levels of dysregulation (Fig. 5*b*). 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 respective brain regions (Fig. 5*c* and Supplementary Table 5). The turquoise modules of the three brain regions shared many editing sites (Fig. 5*d*). 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. 5*e* and Supplementary Table 8). They exhibited significant cortex- and cerebellum-specific expression patterns (Fig. 5*f*), suggesting that the region-specific differential editing may be explained by higher expression in these respective brain regions. These region-specific genes probably have distinct functional roles in ASD.

We also examined 59 editing sites conserved across multiple phylogenetic taxa; these sites probably serve as functionally paramount RNA editing sites in humans³⁸. Thirteen were identified as differentially edited in at least one brain region. Notably, they all exhibited hypoediting in ASD, and six of them were recoding sites (Fig. 5*g*). 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. 5*g*); this site codes for a brain-specific splicing factor that reportedly may cause downregulated splicing in ASD³⁹. This recoding site (S363G) stabilizes protein half-life of NOVA1 (ref. ³⁹), suggesting that the downregulated editing may be an upstream causal factor of downregulated splicing in ASD⁹. Overall these findings strengthen the association between RNA editing and aberrant synaptic signaling in ASD.

Common and brainregion-specific 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. 5*c*, 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* (*ADAR1*, *ADAR2* and *ADAR3*) and Fragile X-related genes with the first principal component of all differential editing sites (Supplementary Fig. 28) also recapitulated many of the turquoise module associations: *FMR1* was significantly associated with the first principal component in both frontal cortex and cerebellum, and *FXR1* was negatively correlated in all three 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 first principal component 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. 5*h,i*). Lastly, *FXR2*, though not associated with the turquoise module in frontal cortex, was significantly positively correlated with the turquoise module in temporal cortex (Fig. 5*c*) and with the principal component 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 in dup15q subjects. Duplication of chromosome 15q11.2-q13.1 (that is, dup15q), which accounts for 0.25–3% ASD diagnoses⁴⁰, clinically manifests with more severe motor impairments and intellectual disability than idiopathic ASD^{40,41}, along with greater magnitude and homogeneous

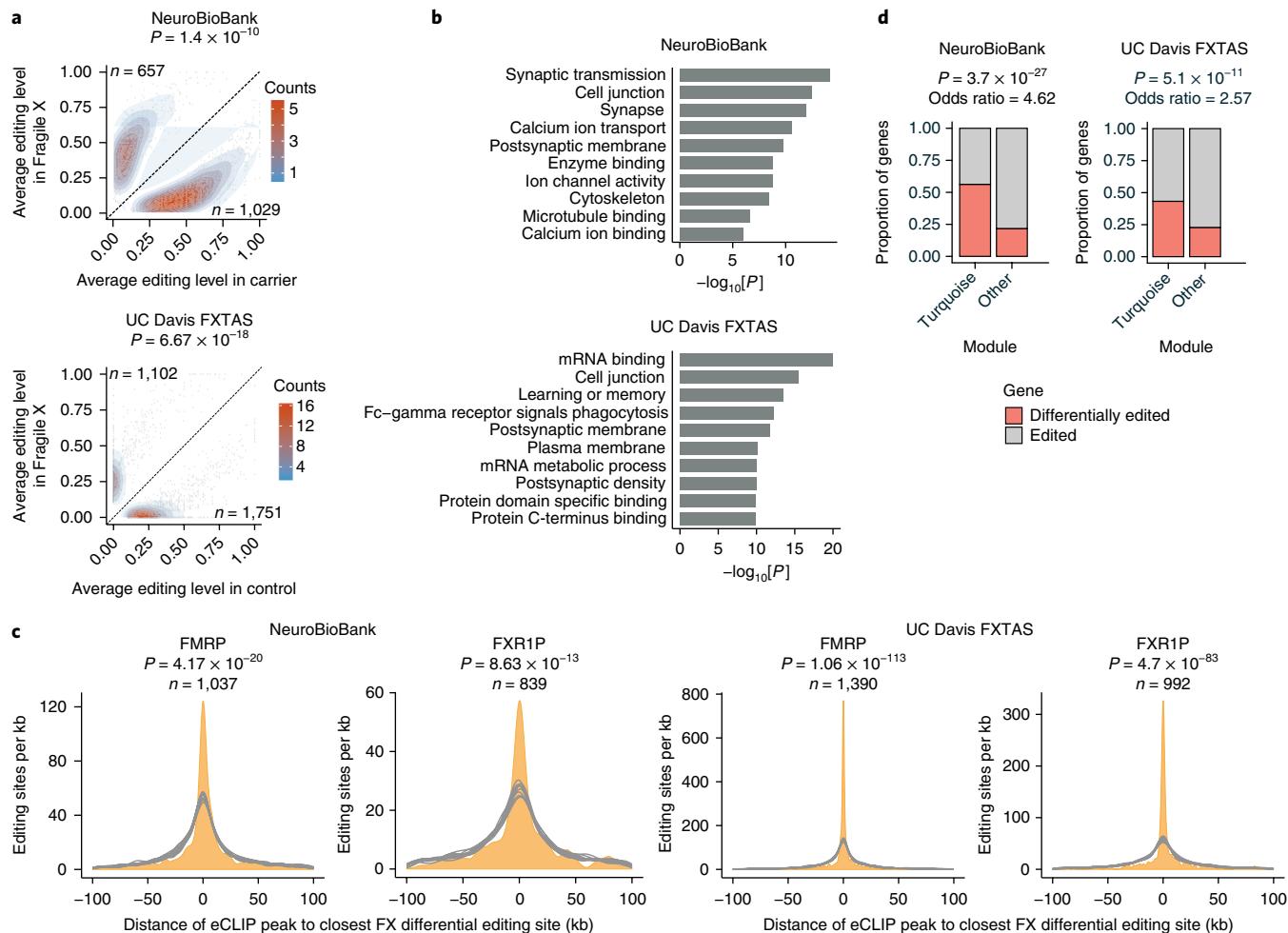


Fig. 4 | Transcriptome-wide differential editing in the frontal cortex of Fragile X subjects and controls. **a**, Average editing levels of differential editing sites in Fragile X subjects compared to carriers (left, NeuroBioBank data set) or controls (right, UC Davis FXTAS data set). Number of editing sites (n) that were up- or downregulated in the subjects are shown; these were compared via chi-squared test (P value shown above plot). **b**, Gene ontology enrichment of genes harboring differential editing sites ($n = 961$ and $1,914$ genes for Neurobiobank and UC Davis FXTAS, respectively; one-tailed Gaussian test, see Methods). **c**, Similar to 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 subjects. P value was calculated via two-tailed Fisher's exact test ($n = 4,051$ and $7,915$ genes in Neurobiobank and UC Davis FXTAS, respectively).

dysregulation of gene expression and splicing⁹. We analyzed RNA editing in dup15q from frontal cortex (eight samples), temporal cortex (nine samples) and cerebellum (five samples) in comparison with covariate matched controls (Supplementary Fig. 29). Dup15q subjects exhibited more profound hypoediting (Fig. 6a) than idiopathic ASD (Figs. 1c and 5a). Correlation between differential editing levels and the intelligence 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, in temporal cortex ($R^2 = 0.64$), cerebellum ($R^2 = 0.36$) and, most prominently, frontal cortex ($R^2 = 0.80$), the region considered most strongly associated with cognitive function⁴². 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 were significantly correlated with those in idiopathic ASD (Fig. 6b), 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 three brain regions of ASD but also reveal an exacerbated hypoediting bias in a subset of ASD subjects 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 subjects across different brain regions and different subject 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, *FMR1* has been subject to a myriad of ASD studies: (1) genes with

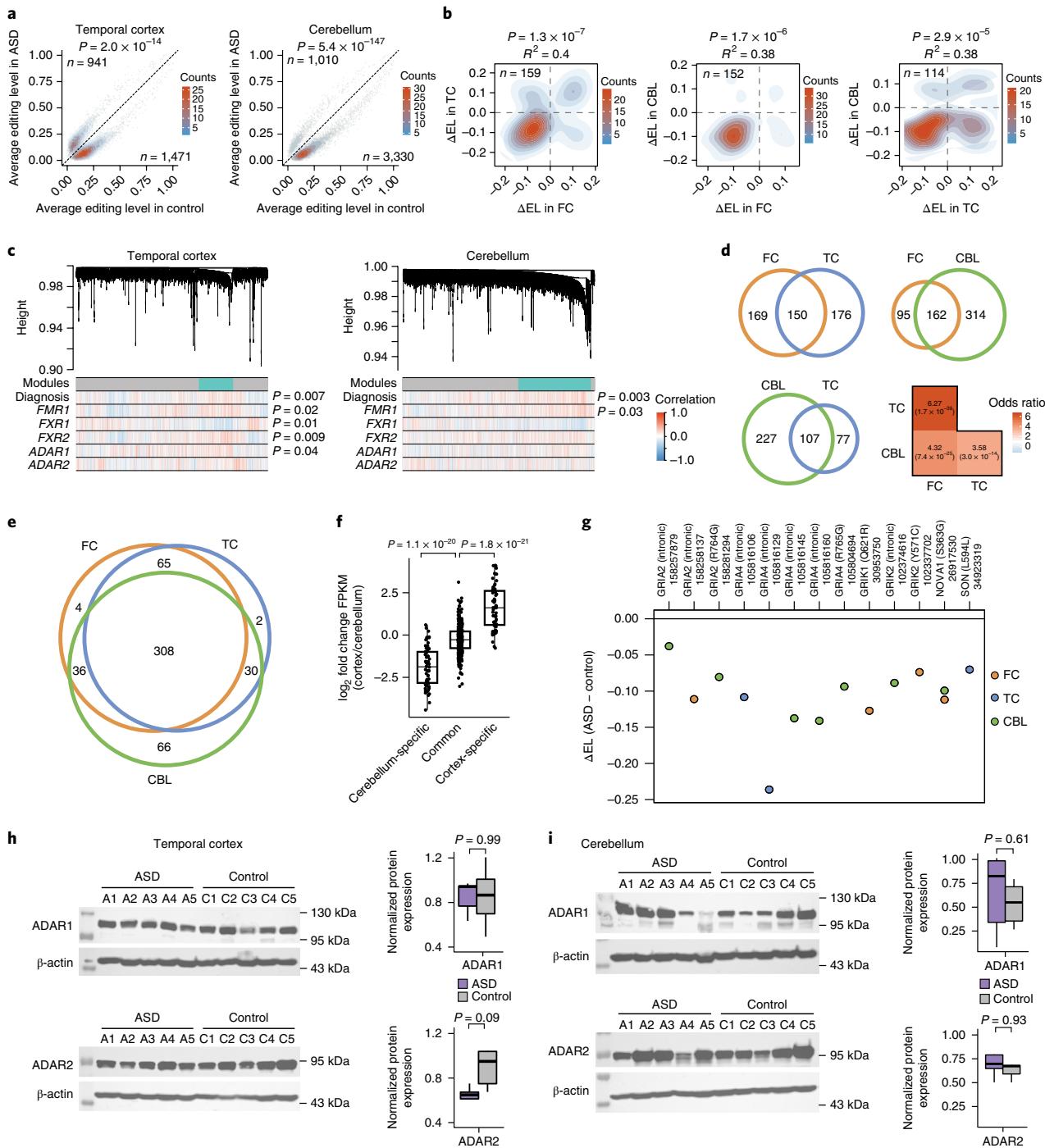


Fig. 5 | RNA editing dysregulation in different brain regions. **a**, Average editing levels of differential editing sites in ASD and controls. Similar to Fig. 1c, but data for temporal cortex and cerebellum are shown, respectively. n , number of differential editing sites; P value calculated by chi-squared 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). TC, temporal cortex; FC, frontal cortex; CBL, cerebellum. Pearson's correlation P and R^2 values are shown. **c**, Similar to Fig. 2d, but 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 parentheses) 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 (\log_2 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 to Fig. 2b, but showing 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 to **h**, but showing 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; minimum and maximum extend to observations at most $1.5 \times IQR$.

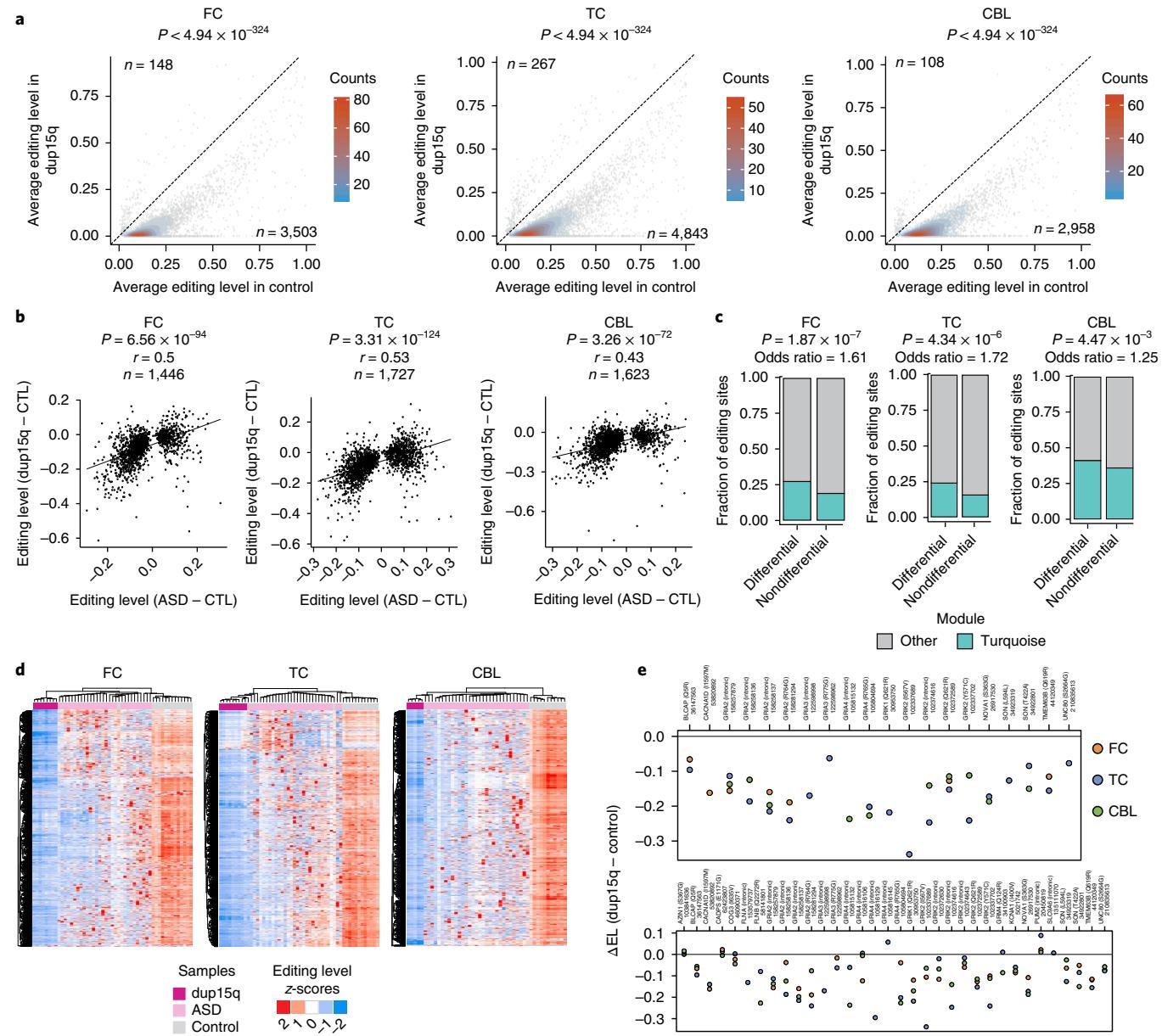


Fig. 6 | Hypoediting in three brain regions of dup15q subjects. **a**, Contour scatterplot of differential editing sites in three brain regions of dup15q subjects vs. matched controls, similar to Fig. 1c. n , number of differential editing sites hypo- and hyperedited. P value calculated using chi-squared test. **b**, Comparison of editing level difference in idiopathic ASD and dup15q subjects relative to their respective controls (CTL). Pearson correlation P values and correlation coefficients (r) are shown. n represents the number of editing sites that are differential in idiopathic ASD and testable (see Methods) in the dup15q subjects. **c**, Overlap between differential editing sites in dup15q subjects and the turquoise modules of the respective brain regions of idiopathic ASD. P values were calculated via two-tailed Fisher's exact test, $n = 3,411, 2,224$ and $3,834$ in FC, TC and CBL, respectively. **d**, Heatmaps (similar to Fig. 1d) of editing sites in **b**, including dup15q subjects, 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: differential sites in at least one brain region of dup15q subjects. Bottom: all testable editing sites.

rare de novo mutations¹⁷, common variation⁴³ and copy number variants⁴⁴ in ASD are enriched in FMRP target genes³⁰; (2) multiple transcriptome analyses have 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 syndrome³⁶ and (4) the protein expression of FMRP is downregulated in ASD subjects⁴⁵. 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 support a model in which FMRP directly mediates the interaction between ADAR and the RNA substrates to promote editing; this advances previous studies of FMRP and RNA editing in mouse, *Drosophila melanogaster* and zebrafish²⁹. The involvement of FXR1P in RNA editing regulation was unknown, and, notably, we observed that FXR1P, probably 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 reveals substantial similarities and highly reproducible patterns in global editing changes in ASD across the three brain regions we profiled, indicating that ASD may affect molecular pathways in general neurological function. Nevertheless, our data also allude to some region-specific editing regulation, such as a down-regulation 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; this 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 neurotypical individuals⁴⁷. Our analyses, although based on a small data set, showed a high correlation between differential editing and IQ scores in all three brain regions. Additionally, dup15q subjects, who are generally known to manifest more severe motor impairments and intellectual disability than those with idiopathic ASD, showed nearly unidirectional and greater severity of hypoediting than idiopathic subjects in all three 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* and *CNTNAP4*), neurexins (*NRXN1* and *NRXN3*), ankyrins (*ANK2*) and neuronal splicing factors (*NOVA1* and *RBFQX1*), which all harbor genetic mutations associated with ASD²⁵. Although causality here is indeterminable, the occurrence of aberrant RNA hypoediting in known ASD risk genes suggests that 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 development; this 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, possibly indicating a functional relationship. Together, this current work suggests that it is important to further explore the role of RNA editing in ASD pathophysiology to determine whether these changes are causal or reflect homeostatic or dyshomeostatic responses.

Accession Codes. Data generated by this study are available at the Gene Expression Omnibus (GEO) with accession codes GSE107895, GSE107867, GSE117776.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41593-018-0287-x>.

Received: 27 December 2017; Accepted: 8 November 2018;

Published online: 17 December 2018

References

- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders 4th edn, text rev.* (APA Publishing, Washington, DC, 2000).
- Rojas, D. C. The role of glutamate and its receptors in autism and the use of glutamate receptor antagonists in treatment. *J. Neural Transm. (Vienna)* **121**, 891–905 (2014).
- Guo, Y. P. & Commons, K. G. Serotonin neuron abnormalities in the BTBR mouse model of autism. *Autism Res.* **10**, 66–77 (2017).
- Ha, S., Sohn, I. J., Kim, N., Sim, H. J. & Cheon, K. A. Characteristics of brains in autism spectrum disorder: structure, function and connectivity across the lifespan. *Exp. Neurobiol.* **24**, 273–284 (2015).
- Nelson, S. B. & Valakh, V. Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. *Neuron* **87**, 684–698 (2015).
- Marchetto, M. C. et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol. Psychiatry* **22**, 820–835 (2017).
- de la Torre-Ubieta, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016).
- Gupta, S. et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* **5**, 5748 (2014).
- Parikhshak, N. N. et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
- Irimia, M. et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
- Wu, Y. E., Parikhshak, N. N., Belgard, T. G. & Geschwind, D. H. Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nat. Neurosci.* **19**, 1463–1476 (2016).
- Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **17**, 83–96 (2016).
- Behm, M. & Öhman, M. RNA editing: a contributor to neuronal dynamics in the mammalian brain. *Trends Genet.* **32**, 165–175 (2016).
- Slotkin, W. & Nishikura, K. Adenosine-to-inosine RNA editing and human disease. *Genome Med.* **5**, 105 (2013).
- Khermesh, K. et al. Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA* **22**, 290–302 (2016).
- Eran, A. et al. Comparative RNA editing in autistic and neurotypical cerebella. *Mol. Psychiatry* **18**, 1041–1048 (2013).
- Iossifov, I. et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
- Parikhshak, N. N. et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
- Bahn, J. H. et al. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–150 (2012).
- Porath, H. T., Carmi, S. & Levanon, E. Y. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.* **5**, 4726 (2014).
- Picardi, E., D'Ercchia, A. M., Lo Giudice, C. & Pesole, G. REDIport: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **45**,(D1), D750–D757 (2017).
- Levanon, E. Y. et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.* **22**, 1001–1005 (2004).
- Oakes, E., Anderson, A., Cohen-Gadol, A. & Hundley, H. A. Adenosine deaminase that acts on RNA 3 (ADAR3) binding to glutamate receptor subunit B pre-mRNA inhibits RNA editing in glioblastoma. *J. Biol. Chem.* **292**, 4326–4335 (2017).
- Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
- Abrahams, B. S. et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).
- Hwang, T. et al. Dynamic regulation of RNA editing in human brain development and disease. *Nat. Neurosci.* **19**, 1093–1099 (2016).
- Liu, X. et al. Disruption of an evolutionarily novel synaptic expression pattern in autism. *PLoS Biol.* **14**, e1002558 (2016).
- Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
- Davis, J. K. & Broadie, K. Multifarious functions of the Fragile X mental retardation protein. *Trends Genet.* **33**, 703–714 (2017).
- Darnell, J. C. et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
- Zhang, Y. et al. The Fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. *EMBO J.* **14**, 5358–5366 (1995).
- Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
- Ascano, M. et al. FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382–386 (2012).
- Vasudevan, S. & Steitz, J. A. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* **128**, 1105–1118 (2007).
- Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA binding proteins. Preprint at *biorXiv* <https://www.biorxiv.org/content/early/2017/08/23/179648> (2017).
- Hagerman, R., Hoem, G. & Hagerman, P. Fragile X and autism: intertwined at the molecular level leading to targeted treatments. *Mol. Autism* **1**, 12 (2010).

37. Abbeduto, L., McDuffie, A. & Thurman, A. J. The fragile X syndrome-autism comorbidity: what do we really know? *Front. Genet.* **5**, 355 (2014).
38. Pinto, Y., Cohen, H. Y. & Levanon, E. Y. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. *Genome Biol.* **15**, R5 (2014).
39. Irimia, M. et al. Evolutionarily conserved A-to-I editing increases protein stability of the alternative splicing factor Nova1. *RNA Biol.* **9**, 12–21 (2012).
40. DiStefano, C. et al. Identification of a distinct developmental and behavioral profile in children with Dup15q syndrome. *J. Neurodev. Disord.* **8**, 19 (2016).
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).
42. Frith, C. & Dolan, R. The role of the prefrontal cortex in higher cognitive functions. *Brain Res. Cogn. Brain Res.* **5**, 175–181 (1996).
43. Jansen, A. et al. Gene-set analysis shows association between FMRP targets and autism spectrum disorder. *Eur. J. Hum. Genet.* **25**, 863–868 (2017).
44. Pinto, D. et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* **94**, 677–694 (2014).
45. Fatemi, S. H. & Folsom, T. D. Dysregulation of fragile X mental retardation protein and metabotropic glutamate receptor 5 in superior frontal cortex of individuals with autism: a postmortem brain study. *Mol. Autism* **2**, 6 (2011).
46. Patzlaff, N. E., Nemec, K. M., Malone, S. G., Li, Y. & Zhao, X. Fragile X related protein 1 (FXRIP) regulates proliferation of adult neural stem cells. *Hum. Mol. Genet.* **26**, 1340–1352 (2017).
47. Charman, T. et al. IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychol. Med.* **41**, 619–627 (2011).

Acknowledgements

Postmortem brain samples used in this study were obtained from the University of Maryland Brain and Tissue Bank, which is a component of the US National Institutes of Health (NIH) NeuroBioBank. We are grateful to the subjects and families who participate in the tissue donation programs. This work was funded by grants from the

NIH to X.X. (HG009417 and HG006264), G.W.Y. (HG004659, HG009417, HG007005 and MH107367), S.T. (T32HG002536), E.L.V.N. (HG009530), V.M.C. (MH094681) and 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.

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.

Competing interests

G.W.Y. is a cofounder of Locana and Eclipse BioInnovations and member of the scientific advisory boards of Locana, Eclipse BioInnovations and Aquinnah Pharmaceuticals. E.V.N. is a cofounder 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.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41593-018-0287-x>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.H.G. or X.X.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018

Methods

RNA sequencing 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 study⁹. 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 ASD⁴⁸. Note that the dup15q samples were analyzed separately as described below. We confirmed that ASD diagnosis was not confounded by age, batch or other biological and technical variables (Supplementary Fig. 2). The final sample set comprised 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 data set. A total of five, eight and nine RNA-seq data sets of dup15q subjects 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 data set to balance covariates (Supplementary Fig. 29), except for batch and brain bank, as there were nominally significant confounding effects (although they did not pass Bonferroni correction) among batch, brain bank and dup15q diagnosis for this subset of data⁹.

Frontal cortex replication data set. For replication of idiopathic results, we downloaded previously published RNA-seq data that were obtained from frontal cortex of 63 ASD and control subjects³⁷. After balancing confounding variables, 22 ASD and 23 control data sets remained, none of which overlapped with subjects from our original data set.

RNA sequencing data sets of Fragile X subjects and carriers and controls.

Postmortem frontal cortex samples of Fragile X subjects and carriers were obtained from the University of Maryland Brain and Tissue Bank and the University of California 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 the manufacturer's instruction. RNA-seq data were collected on an Illumina HiSeq 2000 sequencer.

RNA sequencing read mapping and RNA editing identification. RNA-seq reads were mapped using RASER¹⁹, 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 shows that previous RNA editing identification methods did not detect editing sites in hyperedited regions because of the existence of many mismatches in the reads²⁰. Our implementation of the hyperediting pipeline closely followed a strategy used in a previous study²⁰. In brief, all adenosines in unmapped reads were converted into guanosines. These reads were aligned to a modified hg19 genome in which adenosines were also substituted by guanosines. Uniquely mapped read pairs were obtained from this alignment step, and previously converted adenosines were reinstated. 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 probably resulted from a sequence error¹⁹. A number of posterior filters were then applied to remove RNA editing sites that were probably caused by technical artifacts in sequencing or read mapping⁵⁰.

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 five total reads among which at least two reads were edited; (2) the editing site should satisfy filter (1) in at least five 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 >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 $P < 0.05$ and an effect size $> 5\%$, because an editing change of approximately this magnitude was sufficient to cause dendrite deficits in mice⁵¹.

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 data set²⁷ was performed similarly. However, because only 4,952 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 tested only editing sites where ≥ 5 dup15q and ≥ 5 control samples had ≥ 5 read coverage (defined as testable sites). 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 RNA Integrity Number (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, showing 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 we expected, there is 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 we 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 regressed against age, sex, batch, RIN, brain bank, seqStatPC1, seqStatPC2 and editing level. seqStatPC1 and seqStatPC2 are the first and second principal components encompassing 99% of variance of technical variables as described in our previous work⁹.

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 study²⁶. The original study classified editing sites into three developmental trajectories (constantly minimally edited sites, perpetually highly edited sites and developmentally increasing sites). We recapitulated the three 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 analysis of variance (ANOVA) on each editing site across the six age groups. We considered editing sites passing false discovery rate <0.05 as developmentally increasing sites. Among the remaining sites, those with median editing level >0.5 were categorized as perpetually highly edited sites, and those with median <0.5 as constantly minimally edited sites. Enrichment of editing sites in ASD within these three developmental clusters was performed using Fisher's exact test.

Annotation of editing sites and heatmap generation. Editing sites were annotated using ANNOVAR³². Heatmaps throughout this study were generated using circlize³³.

Principal component analysis. Principal component analysis (PCA) was conducted on differential editing sites to examine associations between principal components and potential confounding covariates. The R function pcomp was used for this purpose. Missing values in the editing level matrix were imputed using the missMDA package³⁴. The principal components were then correlated against technical and biological covariates such as age and gender (Supplementary Fig. 10). The first principal component was predominantly associated with ASD diagnosis, and was thus used as the principal component for differential editing.

Weighted gene coexpression network analysis. The WGCNA package²⁸ in R was used to find modules of correlated editing sites. In multisample 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) ≥5 reads in ≥90% of samples and (2) 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 strategy¹¹. One hundred bootstraps of the data set were carried out to compute the topological overlap matrix of each resampled network. Coediting modules were obtained using the consensus topological overlap matrix of the 100 bootstraps.

WGCNA offers a dynamic tree-cutting algorithm that enables identification of modules at various dendrogram heights and allows delineation of nested modules⁵⁵. 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 because they were 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 probably 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 recommendation⁵⁶. 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, postmortem interval, brain bank, 5' to 3' RNA bias, AT dropout rate, GC dropout rate, mapped bases in intergenic regions, and uniquely mapped reads. The linear model was fit with backwards selection, and the module was deemed associated with ASD diagnosis if $P \leq 0.05$ for the coefficient of this variable.

We tested whether 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. The mRNA expression levels were values corrected after removal of variability contributed by technical covariates⁹.

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 prechilled with liquid nitrogen, spread out in standard tissue culture plates prechilled to -80°C, and ultraviolet crosslinked twice at 254 nM (400 mJ cm⁻²). Crosslinked tissue (50 mg) was then used for each eCLIP experiment, which was performed as described^{32,57}. For controls, we sampled 2% of the pre-immunoprecipitation (postlysis and fragmentation) sample and prepared libraries identically to the FMRP or FXR1 eCLIP (including the membrane size selection step). These libraries served as "size-matched input" (SMInput) to minimize nonspecific background signal in the identical size range on the membrane as well as any inherent biases in ligations, PCR with reverse transcription (RT-PCR), gel migration and transfer steps.

eCLIP-Seq data were analyzed using CLIPper software³², which 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 two-fold of that in the SMInput sample.

To examine FMPR or FXR1P binding relative to RNA editing sites, we compared the distances from eCLIP peaks to turquoise editing sites with the distances between the peaks and gene-matched random adenosines. Only editing sites residing in genes containing at least one eCLIP peak were considered. The closest distance between an editing site or random adenosine and eCLIP peaks was 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 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. 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 the existence of secondary or tertiary RNA structures.

To identify the motifs enriched in eCLIP peaks, we used two alternative methods: HOMER⁵⁸ and DREME⁵⁹. 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 (with settings -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 a 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 subjects and carriers were analyzed similarly to those of the ASD cohorts. Fisher's exact test was used to identify differentially edited sites using pooled subject and carrier data sets ($P \leq 0.05$ and effect size >5%).

Gene ontology enrichment analysis. Gene ontology terms were downloaded from Ensembl⁶⁰. For each query gene, a random control gene was chosen to match gene expression level and gene length ($\pm 10\%$ relative to that of query gene). Gene ontology 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 gene ontology term, a Gaussian distribution was fit to the number of control genes containing this gene ontology term. The enrichment P value of the gene ontology 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). The mixture was incubated on ice for 15 min. Chloroform was added to the mixture and incubated at room temperature for 10 min. The mixture was centrifuged at 12,000g for 15 min, and the top layer was carefully extracted. An 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) according to the manufacturer'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, 2 µl cDNA (corresponding to one-tenth 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 1 min kb⁻¹, 72°C).

TOPO cloning and clonal sequencing. PCR products were run on 1% agarose gel and visualized under ultraviolet light. The correct size band was cut and digested by Zymoclean Gel DNA Recovery kit (Zymo Research, D4002) according to 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 four quadrants and six white clones were randomly selected from each quadrant (total of 24 clones per subject sample per editing site). Each colony was inoculated overnight in LB containing 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 a guanine (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 nondenaturing lysis buffer, 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 antibody to ADAR1 (Santa Cruz, sc-271854) or FMRP (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 antibody to Flag M2 (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, pH 2.8). We loaded 20 µl 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: antibodies to ADAR1 (Santa Cruz, sc-73408), Flag (sc-807), FMRP (Millipore, MAB2160 and Abcam, ab17722), FXR1P (Bethyl Laboratories, A303-892A) and FXR2 (Sigma-Aldrich, F1554). The horseradish peroxidase (HRP)-linked secondary antibodies were used and the blots were visualized with the ECL kit (GE, RPN2232).

Subcellular fractionation. Cells were fractionated following a published protocol with some modifications⁶¹. 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 hypotonic lysis buffer (HLB) + N buffer (10 mM Tris-HCl, 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, 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, pH 7.6, 300 mM NaCl, 7.5 mM MgCl₂, 0.2 mM EDTA, 1 mM dithiothreitol (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 and 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 manufacturer's instruction (Zymo Research, R2072). Reverse transcription 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. However, 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 min 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 (TRCN000159153) constructs were used. We produced lentiviruses via co-transfection of pCMV-d8.91, pSV-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 consortium⁶². Briefly, viruses were collected from conditioned media after 48 h co-transfection. Lentivirus-containing medium was mixed with the same volume of DMEM containing polybrene (8 µg ml⁻¹), which was used to infect HeLa, SK-N-BE(2) and KELLY cells. After 24 h, cells were incubated with puromycin (2 µg ml⁻¹ for HeLa and 1 µg ml⁻¹ for SK-N-BE(2) and KELLY) for 3–7 d. Knockdown efficiency was evaluated by western blot. Cells were lysed

in radioimmunoprecipitation assay (RIPA) buffer containing complete protease inhibitor cocktail. Cell lysates were then resolved through 8% SDS-PAGE and probed by antibodies to ADAR1 (Santa Cruz, sc-271854), ADAR2 (Santa Cruz, sc-73409), FMRP (Millipore, MAB2160), FXR1P (Bethyl Laboratories, A303-892A) and FXR2 (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). The mixture was then incubated on ice for 30 min, sonicated and spun down. Crude protein concentration was obtained using Pierce BCA Protein Assay kit (Thermo Fisher Scientific, 23225). An equal amount of protein was separated using 8% SDS-PAGE and then transferred onto nitrocellulose membrane. The membrane was blocked with 5% nonfat milk (Genesee Scientific, 20-241) and 0.1% Tween 20 in Tris-buffered saline (TBS). The blot was incubated in primary antibody solution against the protein of interest with 5% nonfat milk and 0.1% Tween 20 in TBS overnight at 4 °C on shaker. Antibodies used in this experiment include antibodies to ADAR1 (Santa Cruz, sc-271854), ADAR2 (Santa Cruz, sc-73409), ADAR3 (Santa Cruz, sc-73410) and FMRP (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. β-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-PCR. RNA immunoprecipitation (RIP) was performed according to published protocols with some modifications⁶³. Cells were harvested on the second day of minigene transfection in RIP buffer (25 mM Tris-HCl, 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 (Thermo Fisher Scientific, 10777019), sonicated (10 s three times at 1 min intervals) and centrifuged at 13,000 r.p.m. 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 r.p.m. for 10 min at 4 °C. For immunoprecipitation, lysates were incubated with antibody to FXR1P (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 confluence. 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% normal donkey serum (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 63× oil-immersion objective using Zeiss LSM 780 confocal microscope with ZEN 2011 (Black edition) software and postprocessed 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-squared test under the null hypothesis of equal numbers of up- and downregulated editing sites. Significance of gene set and editing set overlap 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 were 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 Life Sciences Reporting Summary).

Sample size selection. No statistical methods were used to predetermine sample sizes, but our samples sizes are similar to those reported earlier^{8,9}.

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

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 in GEO with accession code GSE107895. RNA-seq data of Fragile X subjects, carriers and controls have been deposited in GEO with accession codes GSE107867 (NeuroBiobank data set) and GSE117776 (UC Davis FXTAS data set). Fastq files of RNA-seq from the idiopathic ASD, dup15q and control brains were obtained from our previous study⁹ 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 in GEO (accession [GSE51264](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51264) / [GSE59288](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59288)).

References

48. Finucane, B. M. et al. 15q duplication syndrome and related disorders. in *GeneReviews* (eds Adam, M. P. et al.) (University of Washington, Seattle, 2016).
49. Ahn, J. & Xiao, X. RASER: reads aligner for SNPs and editing sites of RNA. *Bioinformatics* **31**, 3906–3913 (2015).
50. Lee, J. H., Ang, J. K. & Xiao, X. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA* **19**, 725–732 (2013).
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 (1999).
52. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
53. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
54. Josse, J. & Husson, F. missMDA: a package for handling missing values in multivariate data analysis. *J. Stat. Softw.* **70**, 31 (2016).
55. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
56. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54 (2007).
57. Wheeler, E. C., Van Nostrand, E. L. & Yeo, G. W. Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. *Wiley Interdiscip. Rev. RNA* <https://doi.org/10.1002/wrna.1436> (2018).
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 (2010).
59. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**(W1), W39–W49 (2015).
60. Aken, B. L. et al. The Ensembl gene annotation system. *Database (Oxford)* <https://doi.org/10.1093/database/baw093> (2016).
61. Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat. Protoc.* **11**, 413–428 (2016).
62. Sundaraman, B. et al. Resources for the comprehensive discovery of functional RNA elements. *Mol. Cell* **61**, 903–913 (2016).
63. Bahn, J. H. et al. Genomic analysis of ADAR1 binding and its involvement in multiple RNA processing pathways. *Nat. Commun.* **6**, 6355 (2015).

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

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

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

Immunofluorescence data were obtained using a Zeiss LSM 780 confocal microscope with the ZEN 2011 (Black Edition) software.

Data analysis

ImageJ (version 1.50) was used to analyze Western Blot images.
RASER (version v0.521) was used for read mapping. This software is publicly available.
Computer code for analysis of differential editing is included as Supplementary Software.
ANNOVAR (version 2017jun01, publicly available) was used to annotate the genomic location of editing sites.
circlize (version, publicly available) was used to generate heatmaps in this study.
missMDA (version, publicly available) was used to impute missing values in editing levels.
WGCNA (version, publicly available) was used for co-expression/editing network analysis, this previously published package is freely available as an R package.
CLIPper (version 1, publicly available) was used to analyze eCLIP-Seq data sets.
HOMER (version) and DREME (version) were used for motif analyses, both are freely available.

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All sequencing data (eCLIP-Seq of FMRP, FXR1P, RNA-Seq of Fragile-X patients and carriers/controls) are deposited to the Gene Expression Omnibus (GEO) under accession numbers GSE107895, GSE107867 and GSE117776. RNA-Seq data sets of ASD and control brains were obtained from our previous study and are available in the PsychENCODE website (<https://www.synapse.org/#/Synapse:syn4921369/wiki/235539>).

Field-specific reporting

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

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

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

Sample size	No sample size calculation was performed for this study. The RNA-Seq data of ASD and control brains and those of the replication cohorts are public data sets. Sample sizes were predetermined in those studies and were sufficient for the molecular analyses (expression and/or splicing) in those studies. We analyzed all the available data sets. Our highly reproducible and statistically significant results suggest that the sample sizes were sufficient for our study.
Data exclusions	No data was excluded.
Replication	We included additional RNA-seq data sets to reproduce the findings in our initial RNA-Seq analyses. All results were reproducible. Experimental validations were carried out to verify bioinformatic findings.
Randomization	Randomization in data collection was not applicable in this study because we are using public data sets and our goal was to detect differences between disease and control cohorts. In data analysis, we used randomization to obtain background controls in eCLIP analysis, gene ontology analysis, etc.
Blinding	Blinding in data collection and analyses were not applicable in this study because we are using public data sets and our goal was to detect differences between disease and control cohorts, so we needed to know the category (disease or control) of each data set.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

The following primary mouse monoclonal antibodies were used: ADAR1 (Santa Cruz, sc-73408, Lot # F1417, 1:200), ADAR2 (Santa Cruz, sc-73409, Lot # K0917, 1:200), ADAR3 (Santa Cruz, sc-73410, Lot # B2316, 1:200), FMRP (Millipore, MAB2160, Lot # 2984225, 1:500), β-Actin (Santa Cruz, sc-47778, Lot # J2915, 1:500), β-Tubulin (Santa Cruz, sc-23949, Lot # C0718, 1:200), U1-70K (Santa Cruz, sc-390899, Lot # G0616, 1:100), ADAR1 (Santa Cruz, sc-271854, Lot # F1616, 5μg/each IP sample), FMRP (Millipore, MAB2160, Lot # 2984225, 5μg/each IP sample), FLAG

(Sigma, F1804, Lot # SLBT7654, 2.5ug/each IP sample).

The following primary rabbit polyclonal antibodies were used for the western blot:

FMRP (Abcam, ab17722, Lot # GR272723-1, 1:1000), FXR1P (Bethyl Laboratories, A303-892A, 1:2000), FLAG (Sigma, F7425, Lot # 018M4828V, 1:1000). Goat anti-mouse IgG-HRP (Invitrogen, 31430, Lot # SF252846, 1:2000), goat anti-rabbit IgG-HRP (Invitrogen, 31460, Lot # SE251028, 1:2000).

Validation

All antibodies used are commercially available, and all relevant informations are contained in the Online methods. ADAR1 (clone 15.8.6 quality tested by Santa Cruz Biotech), ADAR2 (clone 1.3.1 quality tested by Santa Cruz Biotech), ADAR3 (clone 3.591 quality tested by Santa Cruz Biotech), FMRP (clone 1C3 quality tested by Millipore), β-Actin (clone C4 quality tested by Santa Cruz Biotech), β-Tubulin (clone 2-28-33 quality tested by Santa Cruz Biotech), U1-70K (clone C-3 quality tested by Santa Cruz Biotech), ADAR1 (clone D-8 quality tested by Santa Cruz Biotech), FLAG (clone M2 quality tested by Sigma), FMRP (rabbit polyclone quality tested by Abcam), FXR1P (rabbit polyclone quality tested by Bethyl Laboratories), FLAG (rabbit polyclone quality tested by Sigma), Goat anti-mouse and anti-rabbit IgG-HRP (goat polyclone quality tested by Invitrogen). Antibodies were properly titrated to the right concentration prior to experimental used.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

HeLa, HEK293T, SK-N-BE(2) and KELLY cells were purchased from ATCC.

Authentication

The cell lines were aliquots of cells that were originally purchased from and authenticated by ATCC.

Mycoplasma contamination

The cell line tested negative for mycoplasma using Venor GeM Mycoplasma Detection Kit (Sigma-Aldrich, MP0025)

Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified cell lines were used.