Article

RNA editing underlies genetic risk of common inflammatory diseases

https://doi.org/10.1038/s41586-022-05052-x

Received: 24 March 2021

Accepted: 29 June 2022

Published online: 3 August 2022

Check for updates

Qin Li[1], Michael J. Gloudemans[2,3], Jonathan M. Geisinger[1], Boming Fan[4], François Aguet[5], Tao Sun[1], Gokul Ramaswami[1], Yang I. Li[1,6], Jin-Biao Ma[4], Jonathan K. Pritchard[1,7], Stephen B. Montgomery[1,2,8] & Jin Billy Li[1,8]✉

A major challenge in human genetics is to identify the molecular mechanisms of trait-associated and disease-associated variants. To achieve this, quantitative trait locus (QTL) mapping of genetic variants with intermediate molecular phenotypes such as gene expression and splicing have been widely adopted[1,2]. However, despite successes, the molecular basis for a considerable fraction of trait-associated and disease-associated variants remains unclear[3,4]. Here we show that ADAR-mediated adenosine-to-inosine RNA editing, a post-transcriptional event vital for suppressing cellular double-stranded RNA (dsRNA)-mediated innate immune interferon responses[5–11], is an important potential mechanism underlying genetic variants associated with common inflammatory diseases. We identified and characterized 30,319 *cis*-RNA editing QTLs (edQTLs) across 49 human tissues. These edQTLs were significantly enriched in genome-wide association study signals for autoimmune and immune-mediated diseases. Colocalization analysis of edQTLs with disease risk loci further pinpointed key, putatively immunogenic dsRNAs formed by expected inverted repeat *Alu* elements as well as unexpected, highly over-represented *cis*-natural antisense transcripts. Furthermore, inflammatory disease risk variants, in aggregate, were associated with reduced editing of nearby dsRNAs and induced interferon responses in inflammatory diseases. This unique directional effect agrees with the established mechanism that lack of RNA editing by ADAR1 leads to the specific activation of the dsRNA sensor MDA5 and subsequent interferon responses and inflammation[7–9]. Our findings implicate cellular dsRNA editing and sensing as a previously underappreciated mechanism of common inflammatory diseases.

Genome-wide association studies (GWAS) have led to the discovery of hundreds of thousands of risk variants involved in trait and disease aetiology, but understanding their molecular function remains an ongoing challenge. QTL studies, best exemplified by gene expression QTLs (eQTLs), have been successful in bridging GWAS variants to their molecular mechanisms[1,2]. Alternative splicing QTLs (sQTLs) have further expanded discovery of these mechanisms[12]. However, other post-transcriptional processes, such as RNA editing, remain largely unexplored, despite the increasing appreciation of their important functions in health and disease[10,13,14].

One of the most abundant RNA modifications is adenosine-to-inosine (A-to-I) RNA editing catalysed by adenosine deaminases acting on RNA (ADARs) that bind to dsRNA substrates and convert adenosines to inosines[5]. As inosine is recognized as guanosine, RNA editing events can be accurately identified and quantified by standard RNA sequencing, unlike most other RNA modifications[15]. Previous studies have identified millions of RNA editing sites in humans, more than 99% of which are located in inverted repeat *Alu*s (IR*Alu*s) that form dsRNA substrates[16–18]. Key to editing in mammals are two enzymatically active ADAR proteins, ADAR1 and ADAR2, which have distinct physiological functions in vivo[19]. ADAR1, which is ubiquitously expressed across human tissues, has a critical role in suppressing dsRNA sensing that is mediated by MDA5, a cytosolic sensor of 'non-self' dsRNA[7–9] (Fig. 1a). Mice deficient in ADAR1 editing are embryonic lethal due to elevated innate immune responses indicated by the induction of interferon-stimulated genes (ISGs), but can be rescued to full life span when *Mda5* is knocked out[8]. In humans, *ADAR1* loss-of-function and *MDA5* gain-of-function mutations have been identified in rare autoimmune diseases such as Aicardi–Goutieres syndrome[6,20], further establishing the ADAR1–dsRNA–MDA5 axis as an underlying mechanism in immune disease (Fig. 1a). Protective, loss-of-function alleles in *MDA5* have also been found in GWAS of common inflammatory diseases such as type 1 diabetes[21,22], psoriasis[23], inflammatory bowel disease (IBD)[24], vitiligo[22,25], vitamin B_{12} deficiency anaemia[26], hypothyroidism[22] and coronary artery disease (CAD)[22,26]. Furthermore, aberrant editing has been reported in several common autoimmune diseases, including psoriasis, rheumatoid arthritis, systemic

[1]Department of Genetics, Stanford University, Stanford, CA, USA. [2]Department of Pathology, Stanford University, Stanford, CA, USA. [3]Biomedical Informatics Training Program, Stanford University, Stanford, CA, USA. [4]State Key Laboratory of Genetic Engineering, Department of Biochemistry and Biophysics, School of Life Sciences, Fudan University, Shanghai, China. [5]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA. [7]Department of Biology, Stanford University, Stanford, CA, USA. [8]These authors contributed equally: Stephen B. Montgomery, Jin Billy Li. ✉e-mail: jin.billy.li@stanford.edu

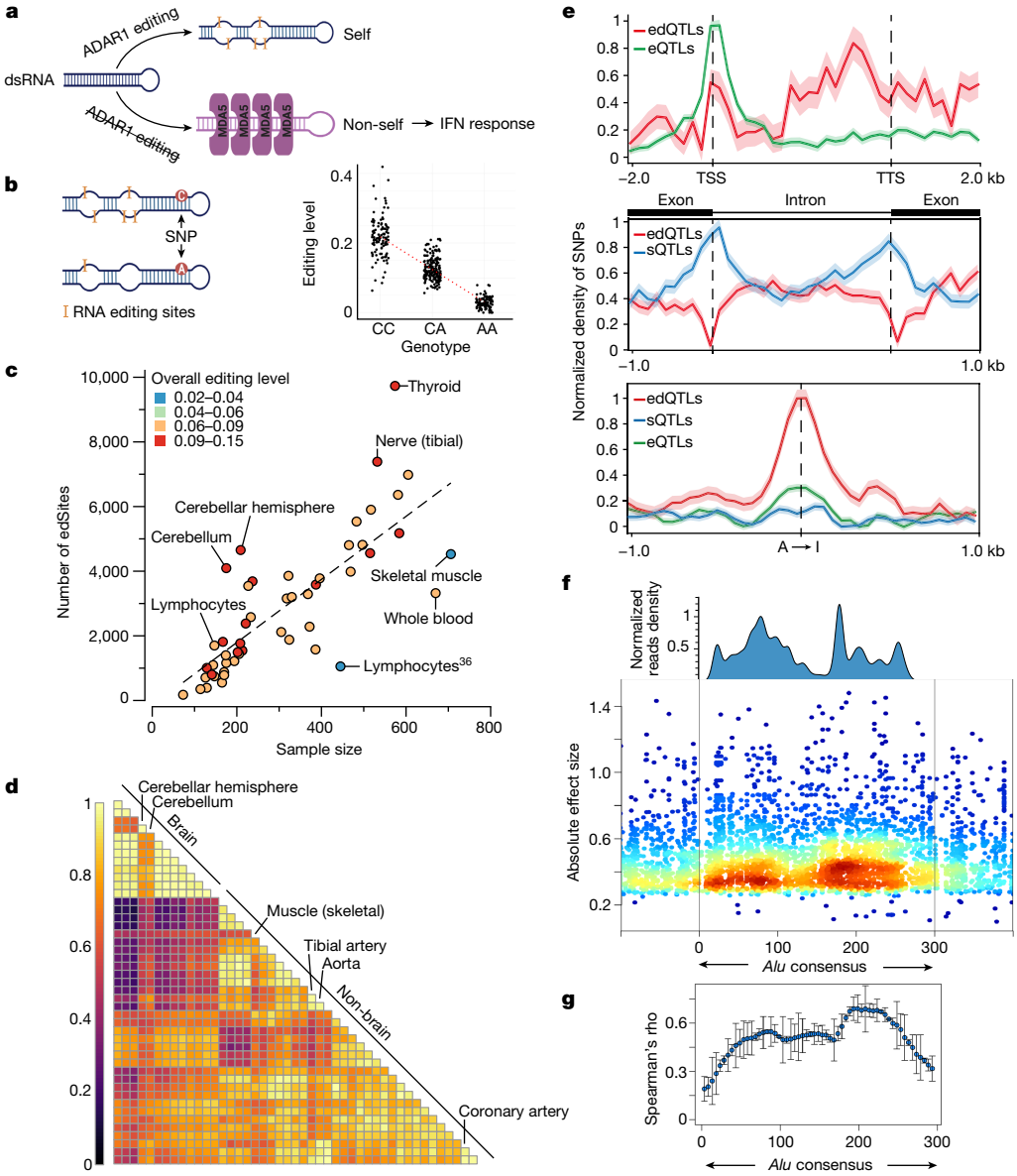


Fig. 1 | RNA edQTL map in human tissues. a, Schematic of the ADAR1–dsRNA–MDA5 axis in innate immunity. **b**, Illustration of the *cis*-edQTL concept. The levels of RNA editing are quantified from bulk RNA sequencing of tissue samples by counting the fraction of edited reads over edited and unedited reads for each editing site, and local genetic effects (edQTLs) are quantified across individuals for each tissue. **c**, The number of edSites mapped in GTEx tissues, compared to the number mapped in lymphoblastoid cell lines (LCLs) from a previous study[36]. Each tissue is colour-coded according to its overall editing level. Tissues of notably high or low overall editing level are labelled. The dashed line represents linear regression between the number of edSites and the sample size across 49 tissues. **d**, Pairwise sharing of edQTLs across tissues. Colours indicate Pearson's correlation coefficient by the magnitude of the edQTL effect size. Tissues with notably distinctive edQTL profiles are labelled. **e**, Comparison of the distribution of SNPs mapped as eQTLs, sQTLs or edQTLs around gene bodies (top), splice junctions (middle) and RNA editing sites (bottom). The density of SNPs is normalized to [0, 1]. The shaded area denotes the standard deviation with the mean as the centre line. TTS, transcription termination site; TSS, transcription start site. **f**, Alignment of ADAR1 binding profile with edQTL SNPs in the *Alu* consensus sequence. The ADAR1 binding profile was estimated using ADAR1 CLIP-seq data (Methods). SNPs are coloured by local density (window size = 10 nt). **g**, Correlation between the binding preference of ADAR1 and the absolute effect sizes of edQTLs. We divided the 300-nt *Alu* consensus sequence into 60 equal bins, and within each bin, we tested the correlation between the average absolute effect sizes of SNPs and the average ADAR1 binding preference using Spearman's method. Error bars represent standard deviation with the mean as the centre point ($n = 22$ autosomes).

lupus erythematosus and multiple sclerosis[27–30]. However, the extent to which common genetic differences in RNA editing may contribute to immune and inflammatory diseases remains to be identified.

Identification of *cis*-RNA edQTLs

In this study, we aimed to obtain a systematic understanding of the roles of A-to-I RNA editing in common and complex traits and diseases. We mapped *cis*-edQTLs (hereafter denoted as edQTLs) using the GTEx V8 RNA sequencing and genotype data (Fig. 1b; Methods), with substantial improvements over previous efforts[31–36] such as larger sample size and more comprehensive RNA editing site annotation. We first measured RNA editing levels—defined as the fraction of edited ('G') transcripts over total ('A' and 'G') transcripts—at single nucleotides ('sites') across a catalogue of more than 2.8 million sites and obtained reliable editing-level quantification for 14,993–60,581 sites per tissue type (Extended Data Fig. 1a,b; Methods). To identify and control for potential confounding factors of editing-level measurement, we performed

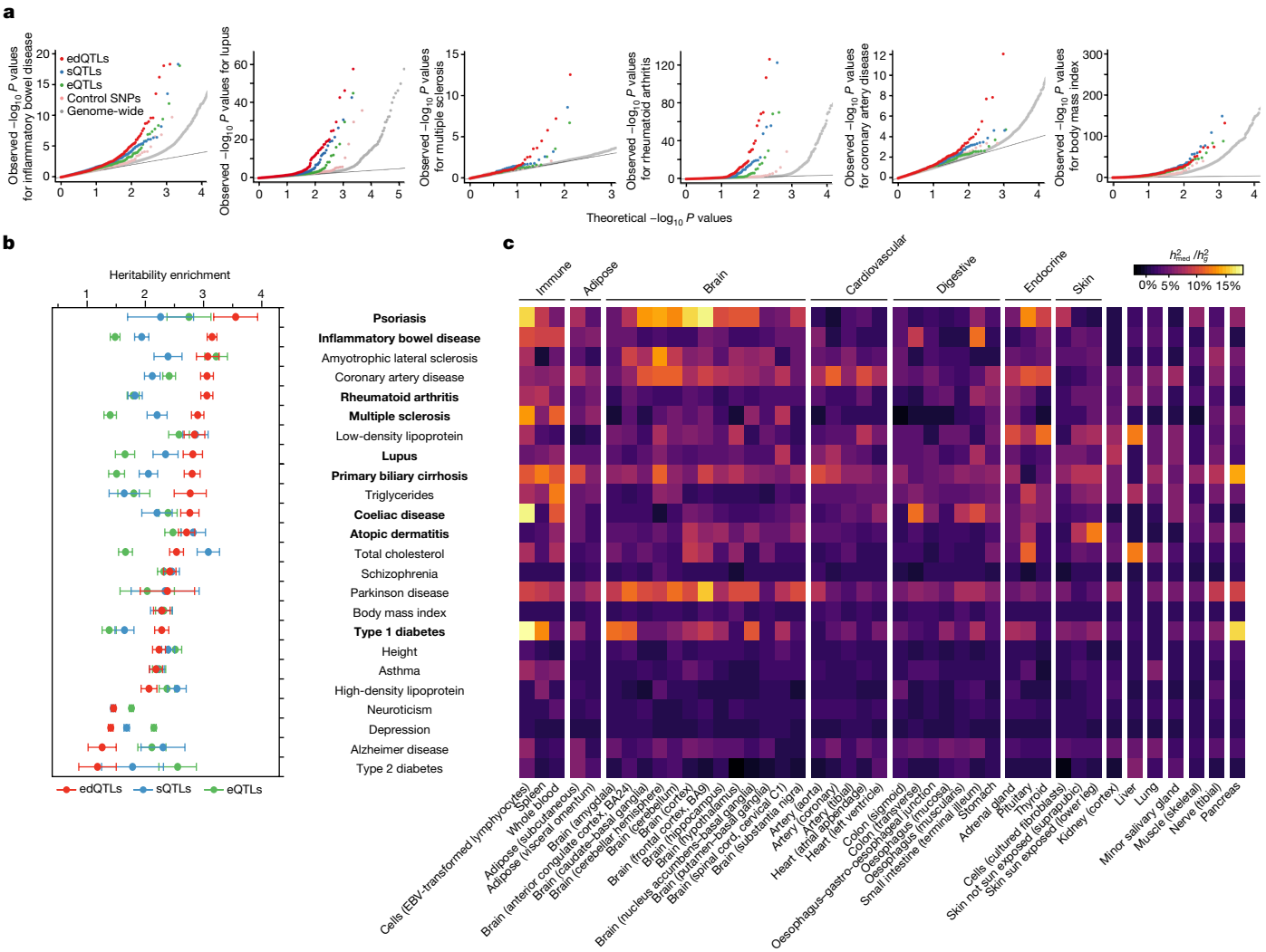


Fig. 2 | Contribution of edQTLs to complex diseases and traits. a, Quantile–quantile plots of *P* values for GWAS SNPs annotated as edQTLs, sQTLs and eQTLs. Random SNPs matching the edQTLs are used as negative controls (Methods). **b**, Enrichment of heritability for 24 human traits and diseases mediated by edQTLs, sQTLs and eQTLs. Autoimmune diseases are shown in bold. Enrichment is measured as the regression coefficient of the mediated expression score regression (Methods). The error bars represent jackknife standard errors around the estimates of enrichment. **c**, Estimated percentage of heritability mediated by edQTLs ($\frac{h^2_{med}}{h^2_g}$) across GTEx tissues for 24 human traits and diseases. Traits and diseases are shown in the same order as in **b**. Tissues with common biological origin are put into tissue groups (top row; see Methods). Sample sizes are provided in Supplementary Table 1.

principal component (PC) analysis and found that the expression level of *ADAR1* correlated with the top PC explaining 8% of the overall variance in editing level (Extended Data Fig. 1c,d; Methods). Moderately (40–60%) edited sites showed high levels of variation across individuals, whereas lowly (less than 10%) or highly (more than 90%) edited sites had little variation (Extended Data Fig. 1e). Similar observations were also made for DNA methylation level[37] and RNA splicing ratios, in which a competition model between splicing isoforms has been proposed to explain how splicing ratios responded to local genetic perturbation[38].

Next, we identified edQTLs in each GTEx tissue type using the QTL mapping pipeline widely adopted by the GTEx Consortium[2,39]. We considered proximal (within ±100 kb of editing sites) common variants (minor allele frequency of more than 5%) and editing sites observed in at least 60 samples of each tissue type (Methods). Of all 287,965 editing sites tested, we identified edQTL for 30,319 sites (10.6%, false discovery rate of less than 5%, permutation-based; Methods); hereafter, we denote editing sites with edQTLs as edSites. In addition, edQTLs were present in 32% (7,165) of all edited genes, which we denote hereafter as edGenes (Extended Data Fig. 2a). Within edGenes, the majority (approximately 60%) of the edQTLs affected multiple editing sites at the same time (Extended Data Fig. 2b), which is expected as editing sites tend to be located in close proximity[14,18]. In addition, nearly 30% of the edGenes had multiple independent edQTLs (Extended Data Fig. 2c), indicating that editing sites can be co-regulated by multiple independent single-nucleotide polymorphisms (SNPs) (Extended Data Fig. 2d). The number of edSites per tissue type was mostly dependent on sample sizes and overall editing levels (defined as transcriptome-wide fraction of edited transcripts over total transcripts) (Fig. 1c). Compared to a recent study using a smaller dataset[34], we identified approximately nine times more edSites (Extended Data Fig. 2e).

Characterization of edQTLs

We compared the effects of edQTLs between tissues by performing a meta-analysis[40] (Methods). We observed that *cis*-genetic effects on RNA editing were highly consistent across tissues (Fig. 1d and Extended Data Fig. 2f–g′). More specifically, the direction of effects across all tissues was consistent for 87.6% of the edQTLs (Fig. 1d and Extended Data Fig. 2f), whereas only 538 edQTLs showed tissue-specific effects (twofold or more difference in the magnitude of effects; Methods)

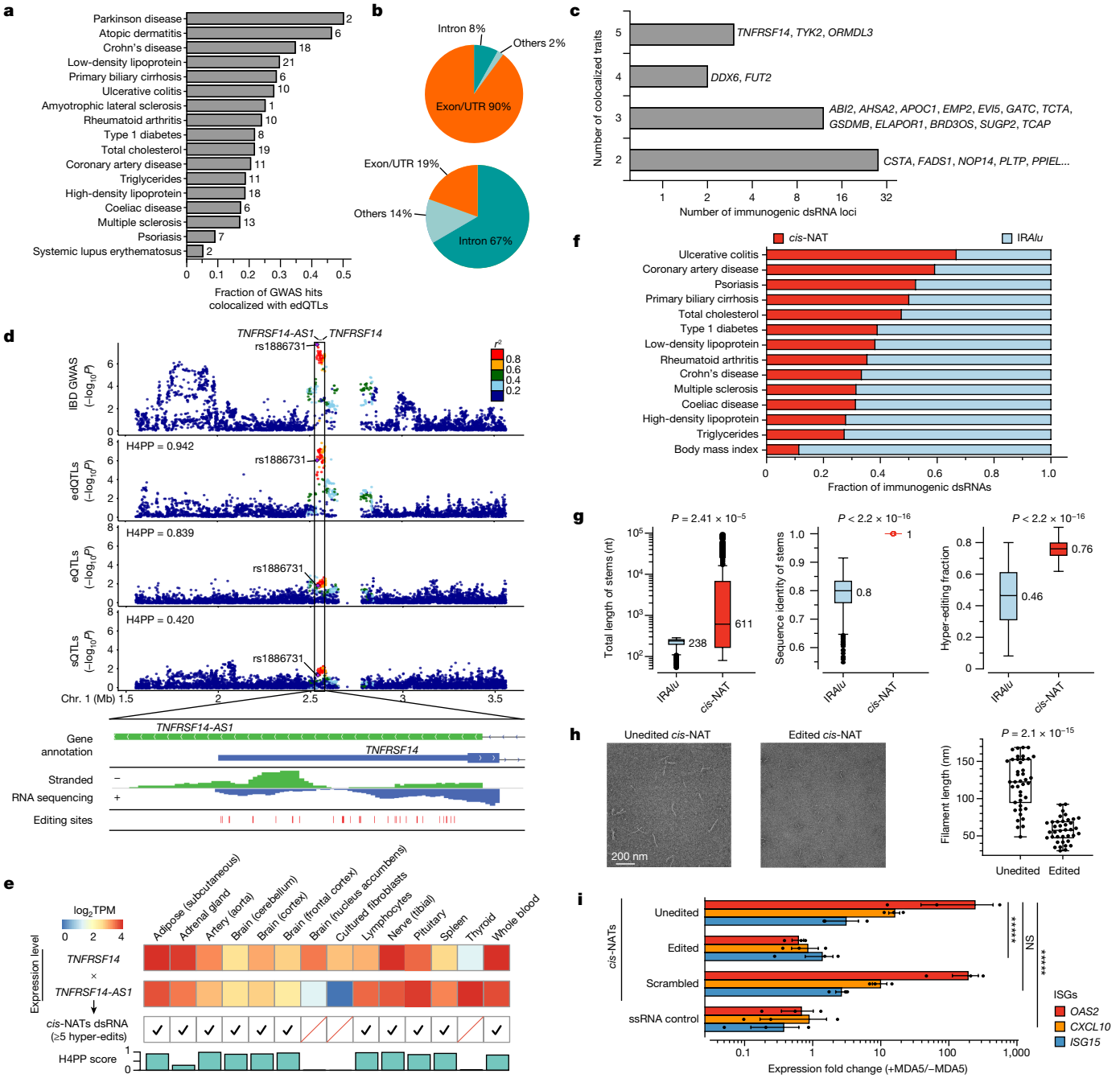


Fig. 3 | Identification, characterization and validation of disease-relevant, putatively immunogenic dsRNAs. a, Fractions of GWAS loci explained by edQTLs. The total number of colocalized loci is shown for each disease. **b**, Genomic distribution of putatively immunogenic dsRNAs (top) and all dsRNAs with edSites (bottom). **c**, Putatively immunogenic dsRNAs colocalized with multiple diseases. **d**, Top dsRNA loci colocalized with IBD GWAS with QTL signals shown below (top). Evidence supporting dsRNA formation by *cis*-NATs shown in the top panel is also displayed (bottom). Bidirectional transcription is supported by gene annotation (top row) and by stranded RNA sequencing (middle row). The formation of dsRNA is supported by hyper-editing (bottom row). H4PP, H4 posterior probability. **e**, Validation of *cis*-NAT dsRNA formation in GTEx tissues. Expression levels are shown for tissues with one or more *cis*-NAT genes expressed. The formation of dsRNA is verified by hyper-editing. H4PP scores are shown for colocalization between edQTL and GWAS. TPM, transcripts per million. **f**, Fraction of IR*Alu* or *cis*-NAT dsRNAs colocalized with immune-related diseases. Data for body mass index are shown as a control. **g**, Comparative analysis of IR*Alu* and *cis*-NAT dsRNAs in three dsRNA features key to MDA5 sensing using two-sided Mann–Whitney *U*-test. **h**, Negative-stain electron microscopy images of the MDA5 filaments along the unedited (left) and edited (middle) *cis*-NAT dsRNAs. Comparison of filament length using two-sided unpaired Student's *t*-test is also shown (right). Points are biologically independent filaments. **i**, Real-time PCR measurements of representative ISGs in *ADAR1*^{E912A} cells, transfected with plasmids expressing *CSTA–PLTP cis*-NAT dsRNA (unedited), *CSTA–PLTP cis*-NAT dsRNA with scrambled sequences (scrambled) or single-stranded RNA (ssRNA control). *cis*-NAT dsRNAs were also expressed in *ADAR1*^{WT} cells (edited). Expression fold changes were calculated by the ratio between cells with and without doxycycline-inducible MDA5 (+MDA5 and −MDA5, respectively). $n = 3$, ****$P < 0.00001$; not significant (NS) $P > 0.5$ (two-sided Wilcoxon signed-rank test). The error bars represent the standard error with the centre as the mean. Box plots in **g** and **h** show interquartile ranges and the median, with whiskers extending to minima and maxima.

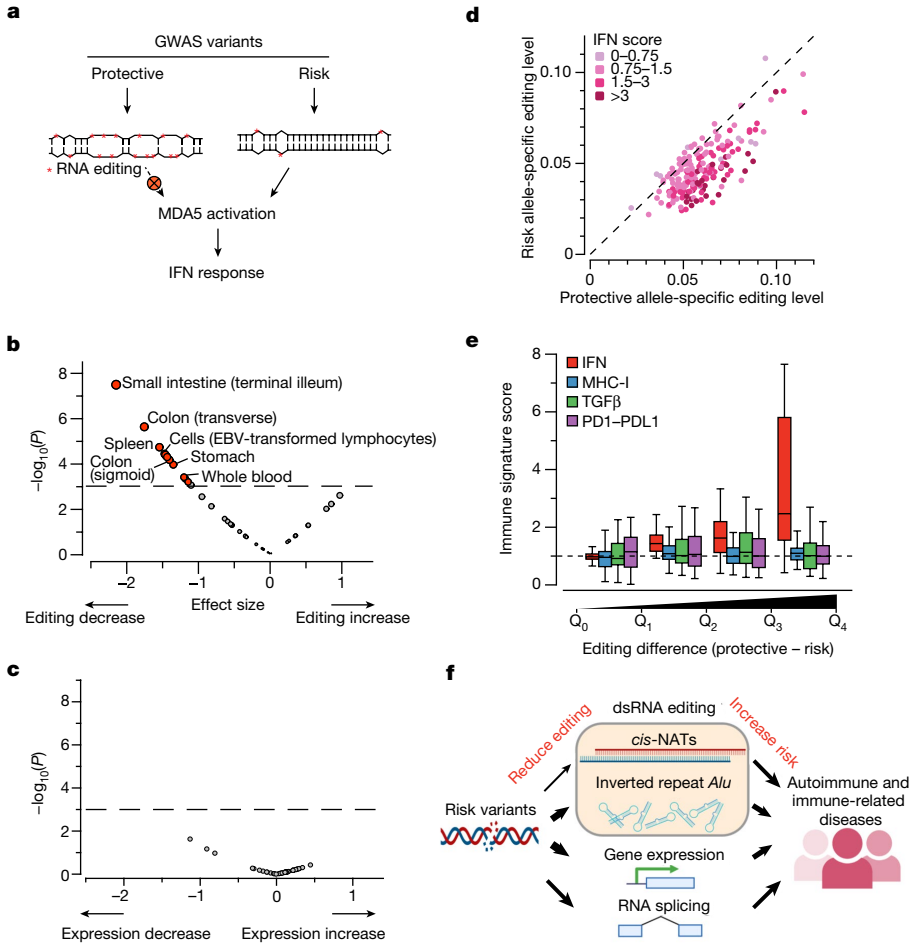


Fig. 4 | Risk variants of inflammatory diseases collectively associated with reduced nearby dsRNA editing levels and induction of interferon response. **a**, Schematic showing that risk variants, by association with reduced editing levels of immunogenic dsRNAs, may lead to MDA5-mediated innate immunity. **b**,**c**, Genome-wide directional effects of risk variants on editing levels (**b**) and expression levels (**c**) observed in IBD. Effect sizes estimated through signed linkage disequilibrium profile regression (Methods) are shown on the *x* axis to denote the overall direction of effects of risk variants on editing level, and are plotted against genome-wide significance (*y* axis) for each GTEx tissue. Tissues showing significant directional effects are coloured in red (Bonferroni-corrected multiple comparison $P < 1 \times 10^{-3}$, shown as a horizontal dashed line). **d**, Risk variants associated with reduced RNA editing levels and induction of ISG expression in rheumatoid arthritis. RNA sequencing data from patient-derived samples ($n = 152$ biologically independent synovial tissues)[59] were used to calculate the overall editing levels of dsRNAs associated with risk versus protective alleles defined by GWAS (see Methods). The diagonal dashed line shows where $y = x$. Each sample is coloured by interferon (IFN) scores calculated using expression levels of signature ISGs from the same RNA sequencing samples (see Methods). **e**, Immune signature scores of the rheumatoid arthritis samples grouped by the extent of reduced editing levels (protective − risk alleles) in quantiles (Q), from low to high. In each group, $n = 38$ biologically independent samples. Expression of predefined signature genes encoding interferon (****$P < 0.0001$, MHC-I (NS, $P \geq 0.05$), TGFβ (NS, $P \geq 0.05$) and PD1–PDL1 (NS, $P \geq 0.05$) immune pathways[49] were used to calculate the scores for each sample. Statistical tests were performed using one-way analysis of variance for the association between each immune signature score and editing-level differences, respectively. Box plots show interquartile ranges and median, with whiskers extending to minima and maxima. Horizontal dashed line indicates baseline immune score of 1. **f**, Schematic illustrating RNA editing of immunogenic dsRNAs bridges the gap between risk variants and susceptibility of autoimmune and immune-related diseases.

(Extended Data Fig. 2h). In addition, 31.1% of edQTLs were found only in one tissue due to tissue-specific gene expression. Our results indicate that the genetic landscape of RNA editing in humans is complex and requires multiple tissue data to fully describe.

We next compared edQTLs with eQTLs and sQTLs identified in the same GTEx dataset[2]. Unlike eQTLs that were enriched near the transcription start sites, edQTLs were enriched in the 3′ untranslated regions (UTRs), which was reflective of the large proportion (43%) of edSites located in the 3′-UTRs (Fig. 1e, top panel). Consistent with a previous report[12], sQTLs, but not edQTLs, were enriched near splicing junctions (Fig. 1e, middle panel). As expected, edQTLs were strongly enriched near editing sites (Fig. 1e, bottom panel). The subtle enrichment observed for eQTLs and sQTLs near editing sites suggests potential overlaps between edQTLs and eQTLs or sQTLs. Indeed, 18.7% and 21.5% of edQTLs were also eQTLs or sQTLs, respectively (Extended Data Fig. 3a; Methods), which is in agreement with a recent study of a smaller scale[34]. In genes with both an edQTL and an eQTL, most of the lead SNPs for edQTL and eQTL were more than 10 kb away from each other and enriched in genomic elements of distinct functional annotations (Extended Data Fig. 3b,c). Our findings highlight that most edQTLs are not detected as significant eQTLs or sQTLs in the GTEx dataset and potentially represent independent genetic regulatory effects.

Our edQTLs map also provides a unique opportunity to investigate how RNA sequence and/or structural changes are associated with editing level in human tissues. We observed that edQTLs in closer proximity to the editing sites generally showed higher statistical significance (Extended Data Fig. 4a), highlighting the effects of proximal regulatory elements such as ADAR-binding sites, RNA sequence motifs and

RNA secondary structures[41]. We subsequently conducted a focused meta-analysis on *Alu* elements where more than 99% of human editing sites reside[15] due to ADAR-binding preference (Fig. 1f; Methods). We observed that the effect size of edQTLs was significantly correlated with the estimated ADAR1-binding strength (Fig. 1g), suggesting that genetic variants could alter editing levels by affecting ADAR1 binding on RNAs. We then evaluated how changes in RNA sequences because of genetic variants may affect the editing levels of nearby sites (Methods). We found that the 'AUAGG' sequence motif centring at the edited 'A' (underlined) was preferential for high editing levels (Extended Data Fig. 4b–d), which agreed with the previously reported 'UAG' motif preferred by ADAR[41]. Furthermore, we observed that edQTL SNPs were 13–54% more enriched in RNA secondary structures recognized by ADAR than nearby non-edQTL SNPs (Extended Data Fig. 4e,f).

Enrichment of edQTLs in GWAS signals

To evaluate the potential role of A-to-I dsRNA editing in common genetic diseases and traits, we assessed the enrichment of edQTLs in GWAS signals of multiple studies. Similar approaches have been successfully applied in eQTL and sQTL studies to reveal major contributions of gene expression and RNA splicing to complex diseases and traits[2,12,42,43]. We found that edQTLs are highly enriched in GWAS signals for autoimmune diseases (as exemplified by IBD, lupus, multiple sclerosis and rheumatoid arthritis) and immune-related diseases (as exemplified by CAD), more than the randomly drawn control SNPs that were matched with a number of SNPs in linkage disequilibrium, allele frequency and gene density (Fig. 2a; Methods). Although eQTLs and sQTLs were also enriched, consistent with previous studies[2,12,43], edQTLs had effects of larger magnitude than either eQTLs and sQTLs (Fig. 2a). This observation still held true when only considering eGenes and sGenes expressed comparably to edGenes, which are generally more highly expressed (Extended Data Fig. 5a). We further confirmed these results by quantitatively assessing the enrichment of QTL-explained heritability for GWAS of complex diseases and traits listed in Supplementary Table 1. In eight out of nine autoimmune diseases tested, edQTLs were more enriched in heritability than eQTLs and sQTLs (Fig. 2b). In addition, edQTLs were enriched in amyotrophic lateral sclerosis, CAD, triglycerides and low-density lipoproteins (Extended Data Fig. 5b), all of which are implicated with immune functions[44–47].

The multi-tissue GTEx data also allowed us to test the tissue-specific contribution of edQTLs in common diseases. Using a recently published approach that specifically aims to distinguish directional, mediated effects from non-directional pleiotropic and linkage effects[48], we estimated the proportion of heritability mediated by edQTLs (defined as $\frac{h^2_{med}}{h^2_g}$; see Methods) in 24 diseases and traits (Supplementary Table 2). Again, an overall higher proportion of heritability was mediated by edQTLs (0.18 ± 0.04) than by eQTLs (0.11 ± 0.02, estimated using the same methods applied to GTEx V8 (ref. [48])) in autoimmune and immune-related diseases, but not in traits without implicated immune contribution (0.033 ± 0.02). Notably, edQTLs of tissues of the immune system (lymphocytes, spleen and whole blood) collectively explained the largest proportion of heritability in most autoimmune and immune-related diseases tested (Fig. 2c and Extended Data Fig. 5c). The edQTL-mediated heritability was also enriched in known tissues of disease relevance, such as digestive tissues (small intestine and colon) for IBD and coeliac disease, brain tissues for amyotrophic lateral sclerosis and Parkinson disease, cardiovascular tissues for CAD, and the pancreas for type 1 diabetes (Fig. 2c and Extended Data Fig. 5c').

In addition to the GWAS of autoimmune and immune-related diseases, we evaluated the edQTL enrichment in GWAS signals of 33 highly heritable immune traits defined in a recent study[49]. We found that edQTLs were more enriched for GWAS of interferon response-related immune traits than other immune traits tested (Extended Data Fig. 6). This is consistent with genetic data from human and mouse showing that lack of ADAR1 RNA editing triggers an MDA5-mediated innate immune response involving interferon[7–9]. Together, our data provide compelling evidence that edQTLs make a significant contribution to the heritability of autoimmune and immune-related diseases, presumably through triggering the interferon response mediated by dsRNA editing and sensing.

Identification of immunogenic dsRNAs

The genome-wide significant enrichment of edQTLs in heritability of autoimmune and immune-related diseases prompted us to pinpoint specific dsRNA loci of disease relevance. We denote these dsRNAs as putatively immunogenic dsRNAs whose sufficient editing is important to suppress autoimmunity, presumably by evading activation of MDA5. We identified putatively immunogenic dsRNAs by systematically investigating signal colocalization between edQTLs found in 49 human tissues and previously reported genetic variants obtained from 24 GWAS, including 17 diseases and traits implicated with immune functions (Supplementary Table 3; Methods). In total, we identified 1,974 colocalization events (loci × edGenes), linking 17 immune-related diseases to 194 genes expressing putatively immunogenic dsRNAs (Fig. 3a and Supplementary Table 3). Because dsRNAs must be sensed by cytosolic MDA5 to elicit immunogenicity, we reasoned that they should be predominantly located in exons or UTRs instead of introns to be present in the cytosol. Indeed, of the 194 putatively immunogenic dsRNAs, 178 (92%) were located in exons, specifically in UTRs where long dsRNA structures are often formed (Fig. 3b, top). By contrast, of all 15,620 dsRNAs associated with edQTLs identified in this work, only 2,967 (19%) were located in exons/UTRs, compared to 10,465 (67%) located in introns where most IR*Alu* dsRNAs reside (Fig. 3b, bottom).

Characterization of immunogenic dsRNAs

We characterized the 194 putatively immunogenic dsRNAs. A majority of them (130, 67%) were located in IR*Alu*s, which appeared to be substantially lower than expected because almost all long dsRNAs are thought to be formed by IR*Alu*s in human[50] (see below). Of the 194 dsRNA, 42 (22%) were shared between at least two diseases (Fig. 3c), suggesting that the immunogenicity of dsRNAs can serve as a common cause of susceptibility for multiple diseases. For example, a top candidate dsRNA found in *TNFRSF14* was shared in five diseases/traits. Instead of being formed by IR*Alu*s, this dsRNA was formed by overlapping genes transcribed in opposite directions, also known as *cis*-natural antisense transcripts (*cis*-NATs)[51] (Fig. 3d, top). Unlike IR*Alu*s that fold into intramolecular dsRNAs, *cis*-NATs can form intermolecular, perfect dsRNAs when not edited. Through a common putative causal SNP, rs1886731, the locus was most strongly shared between the IBD GWAS and edQTL, rather than with an eQTL or sQTL (Fig. 3d, top). The formation of dsRNA by this *cis*-NAT was further supported by three pieces of evidence. First, the overlap of two genes (*TNFRSF14* and *TNFRSF14-AS1*) transcribed in opposite directions was supported by strand-specific RNA sequencing data (Fig. 3d, bottom; Methods). Second, the approximately 500-bp overlapping region was hyper-edited[18], with 62 out of a total of 86 adenosines editable (Fig. 3d, bottom). Third, using multiple GTEx tissues, the expression of both *TNFRSF14* and *TNFRSF14-AS1* was required to observe hyper-editing at the overlapping region and to achieve a significant colocalization score (Fig. 3e).

Next, we expanded our analysis to all 17 immune-related diseases. In 13 of 17 diseases with at least six colocalized loci, we found that a range of 27–59% of all putatively immunogenic dsRNAs were *cis*-NATs, with an average of 33% (Fig. 3f and Supplementary Table 4). By contrast, only 11% of colocalized dsRNAs identified in body mass index were *cis*-NATs. The enrichment of *cis*-NATs as putatively immunogenic dsRNAs in immune-related diseases implies their importance as potential MDA5 ligands with high potency.

We compared dsRNAs formed by IR*Alu*s and *cis*-NATs by their sequence and structural features (Methods). IR*Alu*s typically form 238-bp long dsRNAs (base-pairing region), whereas *cis*-NATs form 611-bp long dsRNAs on average, with the top 25% longer than 6 kb (Fig. 3g, left panel). IR*Alu*s share an average of approximately 80% sequence identity between two *Alu*s, which would result in approximately 48 mismatches along an average 238-bp IR*Alu*. By contrast, *cis*-NATs form perfect dsRNAs because two stems are transcribed from the same genomic locus (Fig. 3g, middle panel). These data suggest that, without being edited, *cis*-NATs, compared to IR*Alus*, generally form longer, better base-paired dsRNAs that are preferred substrates for MDA5 sensing[52], thus making *cis*-NATs potentially more immunogenic. These dsRNAs need to be hyper-edited, presumably by ADAR1, to suppress the immunogenicity. On average, we observed that 46% of reads were hyper-edited in IR*Alu*s, whereas 76% were hyper-edited in *cis*-NATs (Fig. 3g, right panel).

Immunogenicity of *cis*-NAT dsRNAs

To experimentally validate the immunogenicity of dsRNAs formed by *cis*-NATs, we first evaluated whether MDA5 proteins could form filaments on *cis*-NATs in vitro. We tested three *cis*-NAT pairs using negative-stain electron microscopy[53] (Methods). The dsRNA of each *cis*-NAT pair, but not their single-stranded RNA controls, formed MDA5 filaments of varying lengths proportional to the *cis*-NAT dsRNA length (Fig. 3h, left and middle, and Extended Data Fig. 7a–c). Furthermore, the lengths of filaments were significantly reduced when the dsRNAs were in vitro edited by ADAR1 before MDA5 incubation (Fig. 3h and Extended Data Fig. 7d), suggesting that *cis*-NATs need to be edited to evade MDA5 sensing.

We next validated the immunogenicity of *cis*-NATs in human cells with multiple lines of evidence. First, we sought for evidence of dsRNA formation by *cis*-NATs beyond the indicative hyper-editing. By using the publicly available structure mapping[54] and ADAR1-binding data[55] in human cells, we found that of 38 *cis*-NATs expressed in the corresponding cell lines, 29 showed RNA structure signals of high confidence (normalized icSHAPE score of 0.7 or more) for both strands in the overlapping regions, indicating formation of double-stranded structures between the sense and antisense transcripts (Extended Data Fig. 7e,f). Second, we evaluated the ability of *cis*-NAT to induce MDA5-dependent immunogenicity by over-expressing a candidate in human cells. We expressed the *CTSA–PLTP cis*-NAT pair, as a proof of concept, in ADAR1 editing-deficient cells with inducible MDA5 so that the exogenous *cis*-NAT would not be edited to diminish its immunogenicity (Extended Data Fig. 8a; Methods). By transfection of a plasmid that expressed both sense and antisense transcripts at the overlapping region to mimic the formation of *cis*-NATs, we observed elevated immune response as indicated by the induction of three representative ISGs upon expression of MDA5 (Fig. 3i, unedited versus single-stranded RNA control, and Extended Data Fig. 8b–d). Third, we examined the immunogenicity of this *cis*-NAT in cells with wild-type ADAR1. As expected, the immune response was significantly reduced compared to ADAR1 editing-deficient cells (Fig. 3i, unedited versus edited), suggesting that RNA editing by ADAR1 of the *cis*-NAT dsRNAs (Sanger sequencing data shown in Extended Data Fig. 8e) dampened their immunogenicity. Fourth, we hypothesized that the immunogenicity of *cis*-NATs is dictated by the long dsRNA structure rather than the sequences. To test this, we generated a scrambled version of the *cis*-NAT, with the overlapping sequences randomized while maintaining base-pairing complementary (Extended Data Fig. 8b). We found indistinguishable immune induction between the scrambled and the wild-type version of *cis*-NATs (Fig. 3i, unedited versus scrambled), which validates our hypothesis. Together, our analyses indicate that the long dsRNA of *cis*-NATs, without being edited by ADAR1, makes them highly potent ligands for MDA5 to trigger the innate immune responses, echoing their over-representation and importance in inflammatory diseases.

Reduced editing underlies disease risk

Our analyses above suggest functional roles of the putatively immunogenic dsRNAs in autoimmune and immune-related diseases. In the well-established ADAR1–dsRNA–MDA5 mechanism, the activation of MDA5 and the subsequent immune responses are triggered by the lack of ADAR1-mediated editing of immunogenic dsRNAs[7–9]. Very likely, these dsRNAs act in aggregate, rather than alone, to elicit the cellular immunogenicity. Therefore, we reasoned that risk variants of these aforementioned diseases should show directional effects to collectively reduce editing levels of the nearby dsRNAs. Reduced editing of dsRNAs would yield better ligands for the host dsRNA sensor MDA5, thus leading to elevated interferon response (Fig. 4a).

To test this model, we applied signed linkage disequilibrium profile regression[56] to 24 complex traits and diseases (listed in Supplementary Table 1) to assess the direction of genome-wide, collective effects of disease risk variants on editing levels while controlling for linkage disequilibrium structure, allele frequency, gene expression and other potential systematic bias (Methods). Across multiple autoimmune and immune-related diseases, we detected overwhelmingly negative direction of effects (that is, risk GWAS variants are associated with less dsRNA editing in general), which supports our hypothesis that reduced editing levels of dsRNAs collectively may lead to greater disease risk (Fig. 4b and Extended Data Fig. 9). The directional effects on RNA editing were even more significant when tested in tissue types of disease relevance, as exemplified by digestive tissues for IBD, cardiovascular tissues for CAD, pancreatic tissue for type 1 diabetes and brain tissues for Parkinson disease, whereas such directional effects were not observed for eQTLs or control non-directional SNPs (Fig. 4c and Extended Data Fig. 9c).

We further tested the directional effects using RNA sequencing data from patient samples of four immune-related diseases. A challenge of such analysis is that the reduced editing of immunogenic dsRNAs leads to interferon responses, which may subsequently induce expression of ADAR1 and affect the overall editing levels[57,58]. Therefore, the initial reduction of editing associated with risk variants could be masked by the eventual increase of editing in a disease state. To overcome this, we examined allele-specific editing levels, enabling the measurements of editing levels associated with risk versus protective alleles. In total, we analysed 152 synovial tissue samples from patients with rheumatoid arthritis[59], 72 white matter samples from patients with multiple sclerosis[60], 20 peripheral blood mononuclear cell samples from patients with systemic lupus erythematosus[61] and 81 coronary artery samples from patients with CAD[2]. For each disease cohort, we observed significantly reduced editing levels in association with the risk alleles compared with protective alleles (Fig. 4d, relative reduction of $33.7 \pm 15.6\%$, paired Student's t-test, $P < 2.7 \times 10^{-12}$; Extended Data Fig. 10 and Supplementary Table 5). Moreover, the extent of reduction in editing level was positively correlated with elevated interferon response as measured by interferon scores (Methods), but not other immune signatures (Fig. 4e). This finding agrees with the known mechanism of reduced ADAR1-mediated RNA editing resulting in dsRNA-mediated interferon responses[7–9], and provides support for our hypothesis that impaired dsRNA editing contributes to the genetic risk for common autoimmune and immune-related diseases.

Discussion

The main function of dsRNA editing by ADAR1 on cellular transcripts is to evade MDA5-mediated dsRNA sensing and autoimmunity[7–9]. Mutations in ADAR1 and MDA5 can trigger strong dsRNA-mediated immune responses and lead to very rare autoimmune diseases[6,20]. In this work,

we aimed to understand how the editing status of dsRNAs may contribute to common human diseases. We found that common genetic variants associated with RNA editing levels were significantly enriched in GWAS signals of common autoimmune and immune-related diseases and accounted for a significant fraction of disease heritability. We further showed a directional effect that GWAS-associated genetic risk variants, in aggregate, generally are associated with reduced editing levels of nearby dsRNAs, particularly in disease-relevant tissues. The less-edited dsRNAs serve as better substrates of MDA5 to trigger interferon response, which is consistent with the well-established ADAR1–dsRNA–MDA5 axis[7–9] (Fig. 4f). Our findings suggest that the loci at which GWAS risk variants and edQTLs colocalize, by collectively reducing the editing levels of associated dsRNAs (probably in the number of hundreds), contribute to MDA5-dependent interferon response and inflammation. This is further supported by previous findings of MDA5 protective, loss-of-function alleles in GWAS studies for several inflammatory diseases[21,23–26]. Together, our work, built on human genetics and well-established mechanisms, implicates a potential actionable therapeutic approach to treating patients with inflammatory diseases by antagonizing MDA5.

Our analysis also allowed us to pinpoint key, putatively immunogenic dsRNA substrates in relevant tissues and diseases. In addition to the abundant IR*Alu*s as usual suspects of putatively immunogenic dsRNAs, we identified *cis*-NATs as a new species of putatively immunogenic dsRNAs that have been previously overlooked, showcasing the utility of the unbiased analysis empowered by human genetics. We found that *cis*-NATs, although being rare in the transcriptome[62], are highly over-represented as disease-relevant putatively immunogenic dsRNAs. This may be attributed to the dsRNA features that make *cis*-NATs the preferred substrates for MDA5. In summary, our work presents RNA editing as an important mechanism underlying the genetic risk for numerous autoimmune and immune-related diseases, and suggests the manipulation of the dsRNA editing and sensing pathway for potential therapeutic developments.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05052-x.

1. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).
2. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
3. Chun, S. et al. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600–605 (2017).
4. Umans, B. D., Battle, A. & Gilad, Y. Where are the disease-associated eQTLs? *Trends Genet.* **37**, 109–124 (2021).
5. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).
6. Rice, G. I. et al. Mutations in ADAR1 cause Aicardi–Goutieres syndrome associated with a type I interferon signature. *Nat. Genet.* **44**, 1243–1248 (2012).
7. Mannion, N. M. et al. The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.* **9**, 1482–1494 (2014).
8. Liddicoat, B. J. et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115–1120 (2015).
9. Pestal, K. et al. Isoforms of RNA-editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity* **43**, 933–944 (2015).
10. Eisenberg, E. & Levanon, E. Y. A-to-I RNA editing—immune protector and transcriptome diversifier. *Nat. Rev. Genet.* **19**, 473–490 (2018).
11. Samuel, C. E. Adenosine deaminase acting on RNA (ADAR1), a suppressor of double-stranded RNA-triggered innate immune responses. *J. Biol. Chem.* **294**, 1710–1720 (2019).
12. Li, Y. I. et al. RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016).
13. Jonkhout, N. et al. The RNA modification landscape in human disease. *RNA* **23**, 1754–1769 (2017).
14. Walkley, C. R. & Li, J. B. Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. *Genome Biol.* **18**, 205 (2017).
15. Ramaswami, G. & Li, J. B. Identification of human RNA editing sites: a historical perspective. *Methods* **107**, 42–47 (2016).
16. Ramaswami, G. et al. Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods* **9**, 579–581 (2012).
17. Tan, M. H. et al. Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**, 249–254 (2017).
18. Porath, H. T., Knisbacher, B. A., Eisenberg, E. & Levanon, E. Y. Massive A-to-I RNA editing is common across the metazoa and correlates with dsRNA abundance. *Genome Biol.* **18**, 185 (2017).
19. Chalk, A. M., Taylor, S., Heraud-Farlow, J. E. & Walkley, C. R. The majority of A-to-I RNA editing is not required for mammalian homeostasis. *Genome Biol.* **20**, 268 (2019).
20. Rice, G. I. et al. Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nat. Genet.* **46**, 503–509 (2014).
21. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
22. Emdin, C. A. et al. Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat. Commun.* **9**, 1613 (2018).
23. Li, Y. et al. Carriers of rare missense variants in IFIH1 are protected from psoriasis. *J. Invest. Dermatol.* **130**, 2768–2772 (2010).
24. Huang, H. et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178 (2017).
25. Jin, Y., Andersen, G. H. L., Santorico, S. A. & Spritz, R. A. Multiple functional variants of IFIH1, a gene involved in triggering innate immune responses, protect against vitiligo. *J. Invest. Dermatol.* **137**, 522–524 (2017).
26. Sun, B. B. et al. Genetic associations of protein-coding variants in human disease. *Nature* **603**, 95–102 (2022).
27. Shallev, L. et al. Decreased A-to-I RNA editing as a source of keratinocytes' dsRNA in psoriasis. *RNA* **24**, 828–840 (2018).
28. Vlachogiannis, N. I. et al. Increased adenosine-to-inosine RNA editing in rheumatoid arthritis. *J. Autoimmun.* **106**, 102329 (2020).
29. Roth, S. H. et al. Increased RNA editing may provide a source for autoantigens in systemic lupus erythematosus. *Cell Rep.* **23**, 50–57 (2018).
30. Tossberg, J. T., Heinrich, R. M., Farley, V. M., Crooke, P. S. 3rd & Aune, T. M. Adenosine-to-inosine RNA editing of Alu double-stranded (ds)RNAs is markedly decreased in multiple sclerosis and unedited Alu dsRNAs are potent activators of proinflammatory transcriptional responses. *J. Immunol.* **205**, 2606–2617 (2020).
31. Belkadi, A. et al. Identification of genetic variants controlling RNA editing and their effect on RNA structure stabilization. *Eur. J. Hum. Genet.* **28**, 1753–1762 (2020).
32. Ruan, H. et al. GPEdit: the genetic and pharmacogenomic landscape of A-to-I RNA editing in cancers. *Nucleic Acids Res.* **50**, D1231–D1237 (2022).
33. Breen, M. S. et al. Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nat. Neurosci.* **22**, 1402–1412 (2019).
34. Park, E., Jiang, Y., Hao, L., Hui, J. & Xing, Y. Genetic variation and microRNA targeting of A-to-I RNA editing fine tune human tissue transcriptomes. *Genome Biol.* **22**, 77 (2021).
35. Franzen, O. et al. Global analysis of A-to-I RNA editing reveals association with common disease variants. *PeerJ* **6**, e4466 (2018).
36. Park, E. et al. Population and allelic variation of A-to-I RNA editing in human transcriptomes. *Genome Biol.* **18**, 143 (2017).
37. Wijetunga, N. A. et al. The meta-epigenomic structure of purified human stem cell populations is defined at *cis*-regulatory sequences. *Nat. Commun.* **5**, 5195 (2014).
38. Baeza-Centurion, P., Minana, B., Schmiedel, J. M., Valcarcel, J. & Lehner, B. Combinatorial genetics reveals a scaling law for the effects of mutations on splicing. *Cell* **176**, 549–563. e23 (2019).
39. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
40. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).
41. Matthews, M. M. et al. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.* **23**, 426–433 (2016).
42. Hormozdiari, F. et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat. Genet.* **50**, 1041–1047 (2018).
43. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
44. Navab, M., Anantharamaiah, G. M., Reddy, S. T., Van Lenten, B. J. & Fogelman, A. M. HDL as a biomarker, potential therapeutic target, and therapy. *Diabetes* **58**, 2711–2717 (2009).
45. Zeisbrich, M. et al. Hypermetabolic macrophages in rheumatoid arthritis and coronary artery disease due to glycogen synthase kinase 3b inactivation. *Ann. Rheum. Dis.* **77**, 1053–1062 (2018).
46. Misra, M. K., Damotte, V. & Hollenbach, J. A. The immunogenetics of neurological disease. *Immunology* **153**, 399–414 (2018).
47. Suciu, C. F. et al. Oxidized low density lipoproteins: the bridge between atherosclerosis and autoimmunity. Possible implications in accelerated atherosclerosis and for immune intervention in autoimmune rheumatic disorders. *Autoimmun. Rev.* **17**, 366–375 (2018).
48. Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.* **52**, 626–633 (2020).
49. Sayaman, R. W. et al. Germline genetic contribution to the immune landscape of cancer. *Immunity* **54**, 367–386.e8 (2021).
50. Bazak, L. et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **24**, 365–376 (2014).
51. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.* **10**, 637–643 (2009).

52. Dias Junior, A. G., Sampaio, N. G. & Rehwinkel, J. A balancing act: MDA5 in antiviral immunity and autoinflammation. *Trends Microbiol.* **27**, 75–85 (2019).
53. Wu, B. et al. Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. *Cell* **152**, 276–289 (2013).
54. Li, P., Zhou, X., Xu, K. & Zhang, Q. C. RASP: an atlas of transcriptome-wide RNA secondary structure probing data. *Nucleic Acids Res.* **49**, D183–D191 (2021).
55. Zhao, W. et al. POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res.* **50**, D287–D294 (2022).
56. Reshef, Y. A. et al. Detecting genome-wide directional effects of transcription factor binding on polygenic disease risk. *Nat. Genet.* **50**, 1483–1493 (2018).
57. Patterson, J. B. & Samuel, C. E. Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol. Cell. Biol.* **15**, 5376–5388 (1995).
58. Heraud-Farlow, J. E. & Walkley, C. R. The role of RNA editing by ADAR1 in prevention of innate immune sensing of self-RNA. *J. Mol. Med.* **94**, 1095–1102 (2016).
59. Guo, Y. et al. CD40L-dependent pathway is active at various stages of rheumatoid arthritis disease progression. *J. Immunol.* **198**, 4490–4501 (2017).
60. Elkjaer, M. L. et al. Molecular signature of different lesion types in the brain white matter of patients with progressive multiple sclerosis. *Acta Neuropathol. Commun.* **7**, 205 (2019).
61. Tokuyama, M. et al. ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. *Proc. Natl Acad. Sci. USA* **115**, 12565–12572 (2018).
62. Numata, K. et al. Comparative analysis of *cis*-encoded antisense RNAs in eukaryotes. *Gene* **392**, 134–141 (2007).

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Methods

Editing-level quantification in GTEx samples

The GTEx gene expression data used in this study were obtained from the GTEx portal (GTEx Analysis V8 release: https://gtexportal.org/home/datasets) and measured in transcripts per million (TPM). The editing level was quantified on the GTEx Analysis Release V8 (dbGaP accession: phs000424.v8.p2), which consists of a total of 17,382 RNA sequencing (RNA-seq) samples, sequenced in 76-bp long paired-end reads.

We first compiled a list of reference editing sites for quantification. By incorporating known sites in the RADAR database[63], tissue-specific sites identified in GTEx V6p[17], and recently published hyper-editing sites[18], we finalized a list of 2,802,572 human editing sites.

To quantify the editing levels, we computed the ratio of G reads divided by the sum of A and G reads at each site. We included 15,201 RNA-seq samples from 838 donors with matching genotypes in 49 tissues with sample size of 70 or more for editing-level quantification and downstream analyses. For duplicated reads tagged during the RNA-seq mapping process, we chose to retain the read with the highest base quality (if they were the same base quality, a random read was selected). We required that in each tissue, editing sites should be covered by 20 or more non-duplicated reads in 60 or more samples to be considered as testable in the downstream analyses, and the variation across samples is non-zero for edQTL mapping. After applying the above filters, we obtained 14,993–60,581 sites for downstream analyses, varying across tissues (Extended Data Fig. 1a). Codes used for editing-level quantification in GTEx V8 data and the reference editing site list are available at: https://hub.docker.com/r/vanessa/mpileup/.

Because GTEx RNA-seq data are not strand specific, the reads cannot be automatically assigned to sense and antisense strands, so that the quantification of *cis*-NAT editing level may be affected by expression of either of the two overlapping genes. Therefore, we developed a method to measure the RNA editing level of a given site using the reads derived from the corresponding transcript (sense for A-to-G edits and antisense for T-to-C edits), with less influence by the other overlapping transcript. For reads that are edited, they would be derived from the sense transcript if the edits are A-to-G, and from the antisense transcript if the edits are T-to-C. For the rest of the reads that are unedited, we estimated the numbers derived from the sense versus antisense transcripts in proportion to their expression levels.

We validated the accuracy of our method using stranded RNA-seq datasets obtained from ENCODE, which allowed us to count reads in sense and antisense strands separately. We compared editing levels of approximately 800 *cis*-NAT editing sites (20 or more reads coverage) quantified using the estimated reads counts (without taking the strandness into account, using the method described above) versus the exact reads counts (Extended Data Fig. 1b). Overall, the two approaches generated very consistent results of measuring editing levels ($R^2 = 0.98$), thus validating the utility of our method.

cis-edQTL mapping

We first used PC analysis to identify potential confounders in editing-level measurements. Editing-level measurements are usually less confounded than gene expression[17,64]. We found that the top ten PCs collectively contributed to approximately 20% of total variance in editing level (Extended Data Fig. 1c) and PC1 was highly correlated with the expression level of *ADAR1*, agreeing with previous observation[17,33,65,66]. The numbers of PCs regressed out from editing-level measurements (see below) were chosen to maximize the number of detected *cis*-edQTLs with five or less PCs needed in all tissues (we tested 0–10 PCs).

To map edQTLs, we considered all SNPs with minor allele frequency (MAF) of 0.05 or more within ±100 kb of editing sites. Variant call files of genotype data for 838 GTEx donors with matching RNA-seq data were obtained from dbGAP (accession ID: phs000424.v8) based on GRCh38/hg38 reference.

We used FastQTL[39] for edQTL mapping. Raw editing-level measurements were logit-transformed before regressed out PCs, and then normalized to $N(0, 1)$ distribution across individuals within each tissue. The top three genotype PCs together with sex and age were used as covariates for edQTL mapping. For each editing site, the adaptive permutation mode was used with the setting --permute 1000 10000. The beta distribution-extrapolated empirical *P* values from FastQTL were used to calculate trait-level *q* values[67] with a fixed *P* value interval for the estimation of π_0 (lambda = 0.85). A false discovery rate (FDR) threshold of 0.05 or less was applied to identify editing sites with at least one significant edQTL.

To identify the list of all significant variant–site associations for *cis*-edSites, a genome-wide empirical *P* value threshold was defined as the empirical *P* value of the site closest to the 0.05 FDR threshold. The nominal *P* value threshold was then calculated for each editing site based on the beta distribution model (from FastQTL) of the minimum *P* value distribution obtained from the permutations for the gene. For each editing site, variants with a nominal *P* value below the threshold were considered significant and included in the final list of variant–site pairs.

We implemented in the edQTL mapping pipeline with strict removal of gene expression levels to control for potential confounding effects of gene expression. More specifically, for each editing site, we included the expression level of its host gene (measured in TPM values) as an additional covariate alongside genotype and phenotype covariates to be regressed out from the editing-level measurements and used the residuals for edQTL mapping. For *cis*-NAT editing sites, expression levels of both sense and antisense genes were included as two independent covariates.

Tissue sharing of edQTLs

We applied the multivariate adaptive shrinkage implemented in MashR[40] to compare edQTL effect size between tissues. To fit the MashR model, we used the set of approximately 4,000 edQTLs shared between 20 major tissue types to learn the MashR prior, and then fit the MashR model using 40,000 randomly selected variant–trait pairs for the same set of edSites.

We learned data-driven MashR priors by: (1) PC analysis with the number of PCs = 3; (2) empirical covariance of observed *Z*-scores. The data-driven covariances were further denoised by calling cov_ed in MashR. Furthermore, we included the set of canonical covariances, as described in the section '*cis*-edQTL mapping', as an additional MashR prior. We fit the MashR model using the set of randomly selected variant–trait pairs with the error correlation estimated by applying the estimate_null_correlation function in MashR and the priors obtained above. The resulting MashR model was used to compute the posterior mean, standard deviation and local false sign rate for a given variant–trait pair. Estimates of effect size and local false sign rate outputted by MashR were used as metrics of edQTL magnitude and activity, respectively.

Comparative analyses between edQTLs, eQTLs and sQTLs

We obtained eQTLs and sQTLs mapped by the GTEx Analysis Working Group using GTEx V8 data from dbGaP (accession: phs000424.v8.p2). We assessed the sharing of edQTLs with eQTLs and sQTLs by applying Storey's π_1 (ref. [67]). More specifically, we identified significant SNP-editing pairs in a specific tissue, and then used the distribution of the *P* values for these pairs but tested for expression levels or splicing ratios to estimate π_1, the proportion of non-null associations.

For meta-gene analysis, we considered genes that have all three types of QTLs mapped in GTEx tissues. For each gene, the lead SNP for each QTL was used to represent the corresponding *cis*-signal (for edQTLs, the lead SNP of each site was used and compared to the lead SNPs of edQTLs and sQTLs of the same gene). We used the gene-level models

based on the GENCODE[68] V26 transcript annotation, in which isoforms were collapsed to a single 'transcript' per gene as reference to calculate the distribution of QTL SNPs. All genes plus ±2-kb sequences were collapsed to a single meta-gene and further divided into 50 equal bins to calculate the density of SNPs. Splice junctions plus ±1-kb sequences and editing sites plus ±1-kb sequences were treated in the same way for density calculation.

ADAR1 CLIP-seq analysis

Public ADAR1 cross-linking immunoprecipitation and sequencing (CLIP-seq) data in U87MG cells[69] was used for analysis. Standard quality control and filtering were performed using trim_galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to filter for high-quality sequencing reads with adapter sequences removed. Reads shorter than 15 nt after trimming were discarded. To characterize the binding profile of ADAR1 in *Alu* repeats, all reads were first mapped to RefSeq genes using STAR[70] with default settings. We then used BLASTN[71] to align the mapped reads to *Alu* consensus sequences[72] and only kept the best hit for each read (parameters: -evalue 1e-10 -best_hit_score_edge 0.05 -best_hit_overhang 0.25 -perc_identity 50 -strand plus). RNA-seq data of U87MG cells[69] were used as a control set. We processed the control RNA-seq reads in the same way as CLIP-seq reads. Reads passing quality control were mapped to U87MG reference genome sequence using STAR with default settings[70]. The final mapped reads were then aligned to *Alu* consensus sequences as described above. To account for uneven reads coverage in *Alu*, we used the control RNA-seq to calculate the reads density level per base within the *Alu* consensus sequence and then computed the normalization factor per base by dividing the density level by the average density level of the entire *Alu*. Then for CLIP-seq data, read enrichment was calculated by multiplying the read count with the normalization factor.

RNA sequence motif

Given that ADAR recognizes the triplet 'UAG' motif in a position-dependent manner[41,73], we sought to test whether particular type (or types) of nucleotide change at certain positions would have significantly stronger effect on alternating editing levels. For each edSite, we consider all significantly associated SNPs located within 50 nt upstream and 50 nt downstream of the editing site (100 positions in total). We compiled the ribonucleotide changes on RNA according to the SNP and the strand annotation of the gene (for example, if the SNP is a A-to-C mutation on the reverse strand, it will be interpreted as U-to-G on RNA). For each of the 100 positions, we assessed the average effects for all 12 types of ribonucleotide changes (Extended Data Fig. 4c). Of note, not all symmetrical nucleotide changes show the same opposite effects. For example, C-to-A changes at −1 position have an average effect size of −0.8, whereas A-to-C changes at the same position have an average effect size of +0.6. In some cases, the effects were in the same direction, such as A-to-U versus U-to-A changes at +1 position (−0.3 versus −0.2) and G-to-U versus U-to-G changes at +1 position (−0.4 versus −0.1). To simplify the signal and reduce the measurement noise, we further grouped the data to show the final effects of each of the four types of trinucleotides, in regard to the alternative alleles. We plotted the sequence motif using ggseqlogo[74] with the averaged effect size used to adjust for weight. Overall, a strong preference for A and U was observed at −2 and −1 positions, plus preference for Gs at +1 and +2 positions, showing a 'AUAGG' motif.

RNA secondary structure

To understand how mutations affect editing levels through changes on RNA secondary structures, we predicted local RNA structures containing edSites and the associated SNPs. As computational prediction of long RNA molecules is technically challenging[75], we limited the prediction window to ±800 bp around each edSite (in total 1,601 bp) and only considered the SNPs that fall into that window. We further restricted our analysis to the non-*Alu* editing sites. In total, we predicted secondary structures for 8,043 editing sites and subsequently annotated them with structural features using bpRNA[76] (Extended Data Fig. 4e,f). For each of the five structural features (pseudoknots and dangling ends were not considered due to insufficient data), we compared annotations of edVariants to SNPs that are not associated with editing levels but are also found in the same local structure using Mann–Whitney *U*-test.

Enrichment analysis of QTLs in GWAS signal

To make quantile–quantile plots of the GWAS signal annotated with QTL information, we clumped the significant QTLs to obtain independent signals using PLINK[77] (--clump-r2 0.4 --clump-kb 250). To control for overall higher gene expression levels of edGenes than eGenes and sGenes, we matched eGenes and sGenes to edGenes by the median expression levels across tissues. To generate a negative control set, we considered four features to match the control SNP set to edVariants[78]: (1) for MAF distribution, all edVariants were divided to 50 equal bins by allele frequency and the median MAF of each bin was used to select control SNPs of matching allele frequencies from the EUR set of 1000 Genomes; (2) the number of proxy SNPs in linkage disequilibrium (LD; LD 'buddies', $r^2 = 0.7$). Similar to MAF filtering, LD buddies were sampled by matching with the median in each bin of edVariants distribution; (3) for 3′-UTR density, edQTLs are strongly enriched around 3′-UTRs (average enrichment = 2.1, compared to genome-wide), so we matched the number of 3′-UTRs in loci around the control SNPs (enrichment = 2.0), using LD ($r^2 > 0.7$) and physical distance (250 kb) to define the loci; and (4) distance to the nearest transcription termination site (TTS). We sampled the SNP-to-TTS distance (measured from the upstream side of TTS so that control SNPs are located within expressed region) to be within the same deviation estimated from the distribution of edQTL-to-TTS distance.

Heritability assessment and enrichment test in GWAS

We used the mediated expression score regression pipeline for heritability analyses[48]. We first estimated the overall editing scores from individual-level editing level quantified in each GTEx V8 tissue with matched genotype information. Five editing level PCs (described above in the '*cis*-edQTL mapping' section) were used as covariates. For meta-analysis across tissues, the editing scores were generated with edQTL effect sizes estimated using LASSO. Next, we estimated editing-mediated heritability (h2med) using the editing scores from both individual tissues and tissue groups. GWAS summary statistic data of 24 traits described in Supplementary Table 1 were obtained and converted to the .sumstats file format described here https://github.com/bulik/ldsc/wiki/Summary-Statistics-File-Format. LD scores computed from 1000 Genomes phase 3 stratified over a modified version of the baseline LD model v2.0 were downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/. We kept the SNPs known to HapMap 3 as the SNP 'universe'.

To test edQTL enrichment in immune-related traits, we used GWAS from ref. [49]. In total, 33 immune traits were highly heritable among the 139 well-defined immune traits measured from approximately 9,000 patients with cancer enrolled in The Cancer Genome Atlas (TCGA), and GWAS was subsequently performed on these 33 immune traits[49]. The GWAS summary statistics data for these 33 immune traits, including the six interferon response-related immune traits, can be publicly accessed via: https://figshare.com/articles/dataset/Sayaman_et_al_TCGA_Germline-Immune_GWAS_Summary_Statistics/13077920.

Estimating directional effect of RNA editing on complex traits and diseases

We applied the signed LD profile (SLDP) regression method to estimate the directional effects as described in the original paper[56]. We used the 1000 Genomes phase 3 European genotypes to generate the reference

panel for SLDP regression. LD scores of the same population were downloaded for the reference panel https://data.broadinstitute.org/alkesgroup/SLDP/LDscore.tar.gz. We converted the GWAS summary statistic files as described above in the 'Heritability assessment and enrichment test in GWAS' section. We processed the reference panel files by computing a truncated singular value decomposition for each LD block in the reference panel, which were later used to weight the regression conducted by SLDP regression. The signed effect sizes of edQTLs were used as functional annotations to generate signed LD profiles. For the SNPs associated with multiple editing sites, we aggregated the effect sizes across the associated sites within the closest editing cluster using Stouffer's *Z*-score combination method. To explicitly control for the potential signed effects of gene expression, we also used the effect sizes of *cis*-eQTLs from the same set of genes called with edQTLs to generate a separate set of signed LD profiles to be used as a signed background model. As described in the original paper[56], directional effects of minor alleles in five equally sized MAF bins were included in the signed background model to control for systematic signed effects of minor alleles, which could arise from either population stratification or negative selection[79].

We obtained publicly available patient-derived disease samples for allele-specific editing analysis. In total, we tested 152 synovial tissue samples from patients with rheumatoid arthritis (GSE89408), 72 white matter samples from patients with multiple sclerosis (GSE138614), 20 peripheral blood mononuclear cell samples from patients with systemic lupus erythematosus (GSE122459) and 81 coronary artery samples from patients with CAD (GTEx). For each sample, we phased the mapped RNA-seq reads nearby the corresponding disease GWAS risk variants, designating each read as belonging to either the risk or the protective haplotype. We then quantified editing levels of the nearby editing sites (found on the same paired-end reads as the variants or its LD buddies, $r^2 > 0.4$) for each haplotype and computed the overall editing level for sites near risk and protective alleles in each sample.

Interferon scores were calculated with mRNA expression data (measured in TPM) normalized by the median absolute deviation modified *Z*-score. The score was defined as the median absolute deviation modified *Z*-score value of all signature genes in each sample. Interferon signature genes were determined according to a previous study in inflammatory disease[80].

GWAS download and preparation

We downloaded and consistently re-formatted public GWAS summary statistics from various sources (Supplementary Table 1), using tools freely available at https://github.com/mikegloudemans/gwas-download ('download' and 'munge' modules).

SNP selection for colocalization analysis

Because the total number of GWAS traits × GWAS loci × QTL tissues × *cis*-QTL features is very large, it is computationally difficult and unnecessary to run every possible combination. We ran the 'overlap' module given at https://github.com/mikegloudemans/gwas-download to generate a list of all trait–locus–tissue–feature combinations for which the lead GWAS SNP for that trait at the given locus has $P < 5 \times 10^{-8}$, and overlaps an eQTL with $P < 1 \times 10^{-5}$ for the given QTL feature in the given tissue. Each of these combinations represented a single colocalization test to be performed. We performed this process for edQTLs, sQTLs and eQTLs to generate a comprehensive list of 375,000 tests to run (26,000 edQTLs, 183,000 sQTLs and 165,000 eQTLs).

Colocalization analysis

For the set of tests determined in our previous step, we ran colocalization analysis using the tool COLOC[81] with the default parameter settings, estimating allele frequencies from the full set of 1000 Genomes individuals. For each test, we obtained the H4PP; that is, an estimate of the probability that the GWAS and QTL studies share a common causal variant. For subsequent analyses, we considered a test a 'colocalization' if H4PP was more than 0.9, unless otherwise stated. This threshold indicates a high level of support for colocalization. An implementation of the wrapper pipeline that we used for performing COLOC analysis is available at https://github.com/mikegloudemans/ensemble-colocalization-pipeline.

For locuszoom plots shown in Figs. 3 and 4, we used the publicly available R package LocusCompareR to generate plots comparing the signal overlap between GWAS and edQTLs, sQTLs and eQTLs in our locus of interest.

Comparative analyses of dsRNA features for MDA5 sensing

Three dsRNA features key to MDA5 sensing were compared between *cis*-NATs and IR*Alu*s in Fig. 3, including two about RNA secondary structures and one about hyper-editing.

For structural analyses, we used the collection of annotated *cis*-NATs in the human genome[82] together with the newly identified *cis*-NATs from our QTL mapping analysis, in total 501 *cis*-NATs, in comparison to the 1,212 IR*Alu*s annotated in the human genome that are also edited. For each *cis*-NAT, structure prediction was performed on regions where the two transcripts overlap with each other using RNAduplex from the ViennaRNA package[83], whereas for IR*Alu*s, the entire region annotated with IR*Alu*s and editing sites were used for structure prediction using RNAfold also from the ViennaRNA package[83]. To account for the differences of regions not involved in dsRNA formation between *cis*-NATs and IR*Alu*s, we summed the total length of base-pairing regions, excluding mismatches and loops, as the total length of stems (Fig. 3g, left panel), and used the percentage of base-pairing bases in stems, excluding internal loops and multi-loops, as the sequence identity of stems (Fig. 3g, middle panel).

For hyper-editing analysis, we took brain cerebellum samples, where the highest overall editing levels were observed[17]. We used SPRINT[84] to map the RNA-seq reads and call hyper-editing events in the annotated *cis*-NATs and IR*Alu* regions de novo. We required that each dsRNA region had at least five hyper-edits detected to be considered as a hyper-edited dsRNA. Stranded RNA-seq from ENCODE (accession numbers: ENCSR000AEE and ENCSR000AED) was used for STAR mapping[70] to validate the opposite transcription directions of *cis*-NATs.

MDA5 and ADAR1 protein expression and purification

Human MDA5 protein (residues 298 to 1025; Supplementary Table 6) was expressed from pET-50b(+) in *Escherichia coli* C41 cells, induced by adding 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After 20 h of incubation at 18 °C with shaking, cells were harvested by centrifugation, resuspended in a buffer containing 20 mM Tris-HCl, 500 mM NaCl, 5% glycerol, 20 mM imidazole, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), pH 8.0, and lysed with high-pressure homogenization. The proteins were purified to homogeneity using Ni-NTA affinity, cation exchange, second Ni-NTA affinity and size-exclusion chromatography (in that order). HRV3C protease was added after first Ni-NTA affinity chromatography for tagged His6-NusA cleavage.

The human ADAR1p110 isoform (Supplementary Table 6) was expressed in SF9 cells as recombinant protein with a N-terminal Twin-Strep-tag and purified by Strep-affinity chromatography. The cells were suspended and lysed in lysis buffer (20 mM Tris-HCl pH 8.0, 500 M NaCl, 1 mM Tris (2-carboxyethyl) phosphine (TCEP), 0.55% Triton-X100, 1 mM PMSF, protease inhibitors (Sangon Biotech) and 100 ng ml^{-1} RNase A) and purified by Strep-affinity chromatography. The protein was eluted with 50 mM Tris-HCl, 200 mM KCl, 10% glycerol, 1 mM TCEP and 2.5 mM D-desthiobiotin.

Preparation of *cis*-NAT dsRNA and in vitro dsRNA editing

All dsRNAs were in vitro transcribed using T7 RNA polymerase. Two complementary strands were transcripted and purified separately. The pUC19 plasmids containing target sequences were linearized by EcoRI,

extracted with phenol chloroform and precipitated with isopropanol. The in vitro transcription reaction was carried out at 37 °C for 4 h in 100 mM HEPES-K (pH 7.9), 10 mM $MgCl_2$, 10 mM DTT, 6 mM NTP each, 2 mM spermidine, 200 μg ml^{-1} linearized plasmid, 100 μg ml^{-1} T7 RNA polymerase. DNA template was digested with DNase I after reaction. Transcripts were purified by 8% denaturing urea PAGE, extracted from gel slices with 0.3 M sodium acetate and precipitated with isopropanol. RNA of both complementary strands were mixed at a molar ratio of 1:1 in annealing buffer (20 mM Tris-HCl pH 7.5, 50 mM NaCl and 1 mM EDTA) and heated to 90 °C for 3 min then slowly cooled to room temperature. For ADAR1 dsRNA editing in vitro, 0.64 μg annealed dsRNA was diluted and mixed with 0.4 nmol purified hADAR1-p110 to a total volume of 100 μl in buffer containing 50 mM Tris-HCl pH 7.5, 60 mM KCl, 4% glycerol, 0.002% NP-40, 1 mM DTT, 1 mM EDTA, 1 mg ml^{-1} BSA and 0.4 U μl^{-1} recombinant RNase inhibitor (Takara). Reaction was incubated at 37 °C for 60 min and edited RNA was purified using the Absolutely RNA Nanoprep Kit (Agilent).

Negative-staining electron microscopy

Samples including 0.38 μM HsMDA5 (298–1025) and 3.6 ng μl^{-1} dsRNA (regardless of length) were incubated on ice for 60 min with 1 mM ADP•AlF4 in buffer containing 20 mM HEPES, 100 mM NaCl, 2 mM $MgCl_2$, 2 mM DTT and pH 7.5. 5 μl of samples were applied to the glow-discharged 300 mesh carbon-coated copper grids (Beijing Zhongjingkeyi Technology), stained with 0.75% uranyl formate and air-dried. Data were collected on a Talos L120C transmission electron microscope equipped with a 4K × 4K CETA CCD camera (FEI). Images were recorded at a nominal magnification of ×45,000, corresponding to a pixel size of 3.17 Å per pixel. Filament lengths were measured using ImageJ.

Cell lines

HEK293T-ADAR1-E912A-iMDA5-mCherry-pIFN-Lucia cells were maintained in DMEM (11995-065, Gibco) supplemented with 10% FBS (16140-071, Gibco). This cell line was generated by transducing the HEK293T-ADAR1-E912A ADAR1 editing-deficient cell line with three lentiviral vectors encoding the rtTA reverse transactivator, a doxycycline-inducible construct encoding MDA5 linked to mCherry, and the secreted luciferase Lucia (InvivoGen) under the control of an interferon-responsive promoter.

Plasmid construction

pK-mC3-CTSA-PLTP-mR3 was constructed using gene fragments synthesized by Twist Bioscience and standard molecular cloning techniques. This plasmid contains an mClover3 cassette and an mRuby3 cassette facing towards each other with both under the control of separate EF-1α promoters and CAG enhancers. The 3′-UTR of *CTSA* was attached to mClover3 and the 3′-UTR of *PLTP* was attached to mRuby3 to mimic the endogenous overlapping state of these genes in the human genome (Extended Data Fig. 8a). The scrambled version of *cis*-NATs was built on the same backbone as the wild-type *cis*-NAT with the 208-bp overlapping region sequences randomized while maintaining complementarity between sense and antisense strands (Extended Data Fig. 8b). pKER-mClover3 containing an EF-1α promoter-driven and CAG enhancer-driven mClover3 expression cassette with a rabbit β-globin terminator was cloned to the same vector backbone as the *cis*-NAT and used as a ssRNA control.

cis-NAT dsRNA candidate overexpression assay

HEK293T-ADAR1-E912A-iMDA5-mCherry-pIFN-Lucia cells were plated on poly-L-lysine-coated (A-005-C, Millipore Sigma) 24-well plates (353047, BD Falcon) at 100,000 cells per well. Twenty-four hours later, cells were transfected with 500 ng per well of pK-mC3-CTSA-PLTP-mR3, pKER-mClover3 or Lipofectamine alone using Lipofectamine 3000 (TL30001, Thermo Fisher Scientific), according to the manufacturer's instructions. Twenty-four hours post-transfection, media were aspirated and replaced with 500 μl of DMEM containing 0 or 0.1 μg ml^{-1} of doxycycline. Twenty-four hours after the addition of doxycycline, cells were washed with PBS and harvested with 0.5% trypsin. RNA was isolated from cells using the Monarch Total RNA miniprep kit (T2010S, NEB). cDNA was made from 500 ng of RNA using the iScript Advanced cDNA synthesis kit (1725038, Bio-Rad). cDNA was diluted roughly twofold with nuclease-free water and subjected to quantitative PCR using Kapa SYBR FAST 2x qPCR Master Mix and primers derived from Primer Bank (https://pga.mgh.harvard.edu/primerbank/). Primers (200 nM) were used in each reaction and run on the Bio-Rad CFX96 following the manufacturer's instructions for Kapa SYBR Fast. Data were analysed using the $^{\Delta\Delta}$Ct method. Each assay was carried out as three separate biological replicates.

RT–PCR validation of plasmid expression

To determine whether both cassettes on pK-mC3-CTSA-PLTP-mR3 were expressing, isolated RNA was subjected to Turbo DNase (AM2238, Life Technologies) treatment to digest any residual plasmid followed by purification with the Monarch RNA Cleanup Kit (T2040L, NEB) before being subjected to cDNA synthesis with iScript Advanced cDNA synthesis kit. cDNA was diluted roughly twofold with nuclease-free water. Of the diluted cDNA, 1 μl was then amplified via PCR using OneTaq Quick-Load 2X Master Mix with Standard Buffer (M0486S, NEB) and primers designed to amplify from mClover3 through the *CTSA* 3′-UTR or from mRuby3 through the *PLTP* 3′-UTR. The resultant PCR products were subjected to electrophoresis on a 1% agarose gel (Extended Data Fig. 8c).

Reporting summary

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

Data availability

GTEx V8 release editing-level data and edQTL call sets are available on the GTEx Portal: https://gtexportal.org. Details of the GWAS summary statistics used for colocalization are provided in Supplementary Table 1. icSHAPE data were obtained from the RASP database: http://rasp.zhanglab.net/. CLIP data were obtained from the POSTAR3 database: http://postar.ncrnalab.org/.

Code availability

A full pipeline to preprocess the GWAS and edQTL data, prioritize relevant loci, run colocalization tests and generate the associated plots is publicly available at https://github.com/mikegloudemans/rna-editing-coloc and https://github.com/vargasliqin/GTEx_edQTL.

63. Ramaswami, G. & Li, J. B. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* **42**, D109–D113 (2014).
64. Zhang, R. et al. Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nat. Methods* **11**, 51–54 (2014).
65. Ramaswami, G. et al. Genetic mapping uncovers cis-regulatory landscape of RNA editing. *Nat. Commun.* **6**, 8194 (2015).
66. Tran, S. S. et al. Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* **22**, 25–36 (2019).
67. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
68. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
69. Bahn, J. H. et al. Genomic analysis of ADAR1 binding and its involvement in multiple RNA processing pathways. *Nat. Commun.* **6**, 6355 (2015).
70. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
71. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
72. Konkel, M. K. et al. Sequence analysis and characterization of active human Alu subfamilies based on the 1000 Genomes Pilot Project. *Genome Biol. Evol.* **7**, 2608–2622 (2015).
73. Eggington, J. M., Greene, T. & Bass, B. L. Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.* **2**, 319 (2011).

74. Wagih, O. ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).
75. Tian, S. & Das, R. RNA structure through multidimensional chemical mapping. *Q. Rev. Biophys.* **49**, e7 (2016).
76. Danaee, P. et al. bpRNA: large-scale automated annotation and analysis of RNA secondary structure. *Nucleic Acids Res.* **46**, 5381–5394 (2018).
77. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
78. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31**, 418–420 (2015).
79. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
80. Rice, G. I. et al. Assessment of type I interferon signaling in pediatric inflammatory disease. *J. Clin. Immunol.* **37**, 123–132 (2017).
81. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
82. Balbin, O. A. et al. The landscape of antisense gene expression in human cancers. *Genome Res.* **25**, 1068–1079 (2015).
83. Lorenz, R. et al. ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
84. Zhang, F., Lu, Y., Yan, S., Xing, Q. & Tian, W. SPRINT: an SNP-free toolkit for identifying RNA editing sites. *Bioinformatics* **33**, 3538–3548 (2017).

Acknowledgements We acknowledge the GTEx consortium for making the data available for this study; members of the Li and Montgomery laboratories for insightful discussions and suggestions; C. Walkley, J. Engreitz, A. Marderstein and O. de Geode for critical reading of the manuscript; and X. Liu, W. Tsui and V. Sochat for technical support. This work was funded by US National Institutes of Health grants R01 GM102484, R01 GM124215, R01 MH115080, R35 GM144100, R01 AG066490, R01 MH125244, U01 HG009431 and U01 HG007593. Q.L. was partially funded by the American Heart Association Postdoctoral Fellowship (17IFUNP33820059). M.J.G. was funded by NLM training grant T15 LM 007033 and a Stanford Graduate Fellowship.

Author contributions Q.L., M.J.G., S.B.M. and J.B.L. conceptualized the study. Q.L., M.J.G., J.M.G., B.F., G.R., Y.I.L., J.-B.M., S.B.M. and J.B.L. performed the study. Q.L., M.J.G., J.M.G., B.F. and F.A. undertook the formal analysis. Q.L. and J.B.L. wrote the original draft of the manuscript. Q.L., M.J.G., J.M.G., T.S., G.R., Y.I.L., J.K.P., S.B.M. and J.B.L. reviewed and edited the manuscript. J.B.L. acquired funding. S.B.M. and J.B.L. supervised the study.

Competing interests S.B.M. is a consultant for BioMarin, MyOme and Tenaya Therapeutics. J.B.L. is a co-founder of AIRNA Bio and a consultant for Risen Pharma. J.B.L. and Q.L. are named inventors of a provisional patent filed by Stanford University (serial no. 63/473,678), describing a method related to this work. The other authors declare no competing interests.

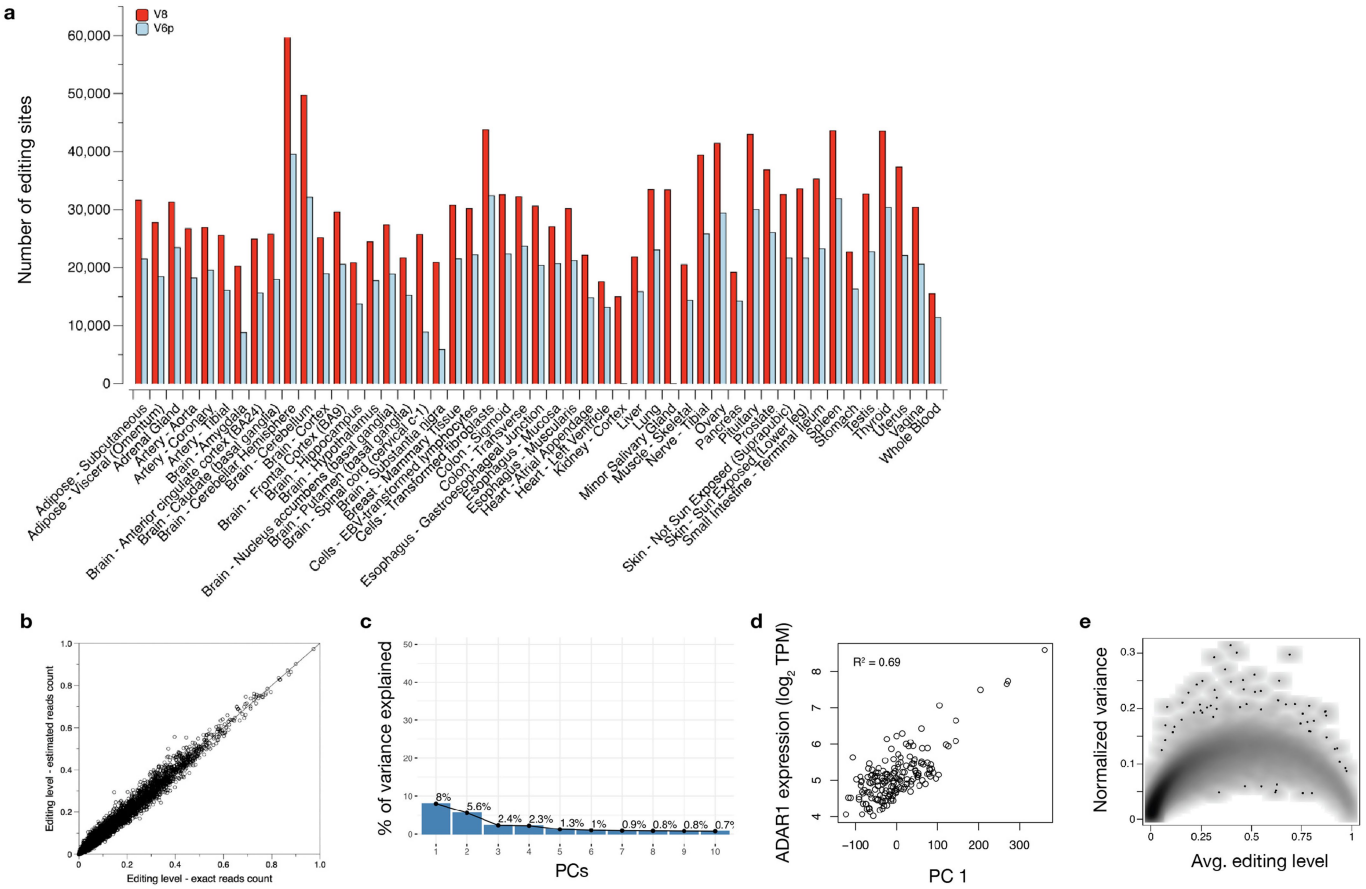
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05052-x.

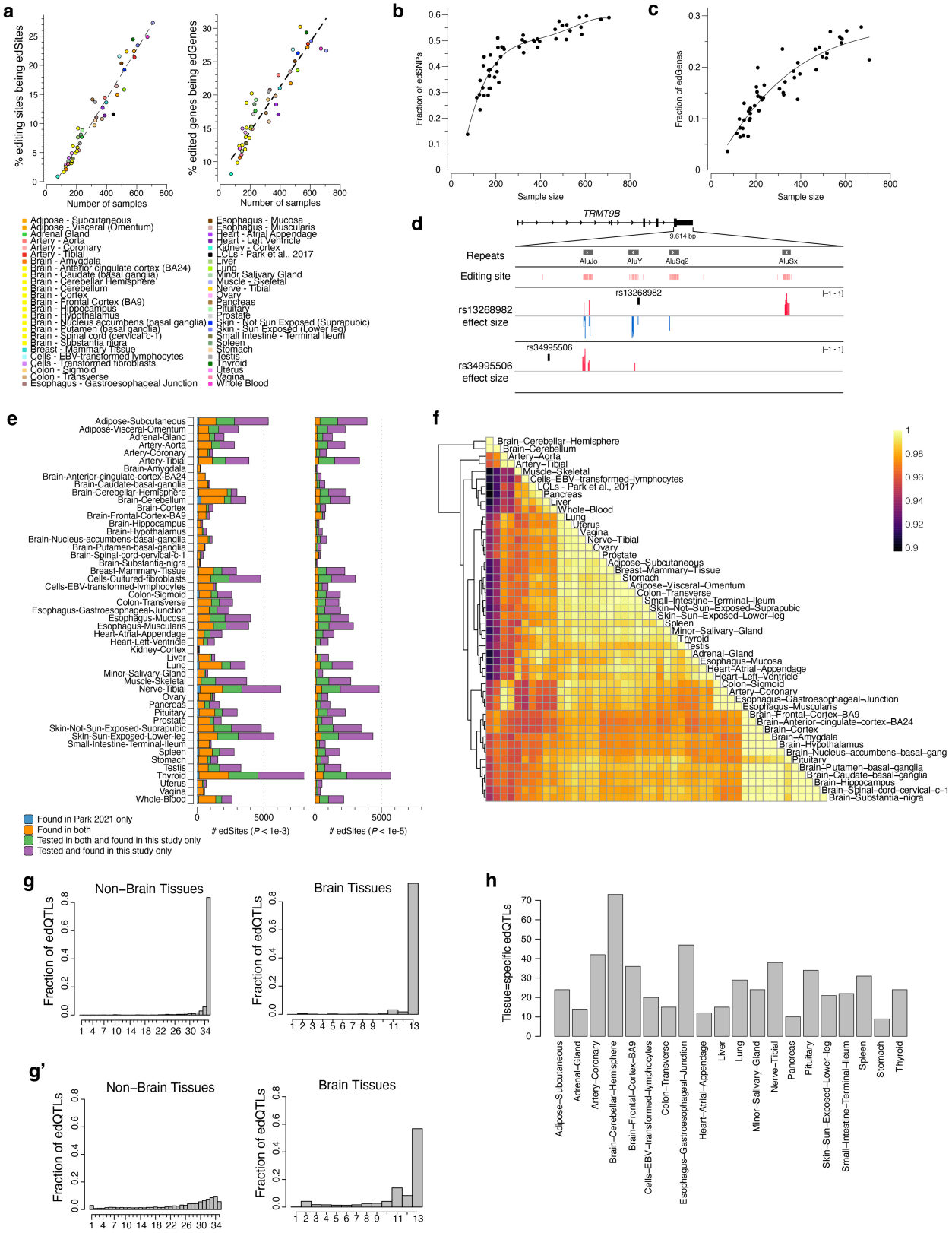
Correspondence and requests for materials should be addressed to Jin Billy Li.

Peer review information *Nature* thanks Kaur Alasoo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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

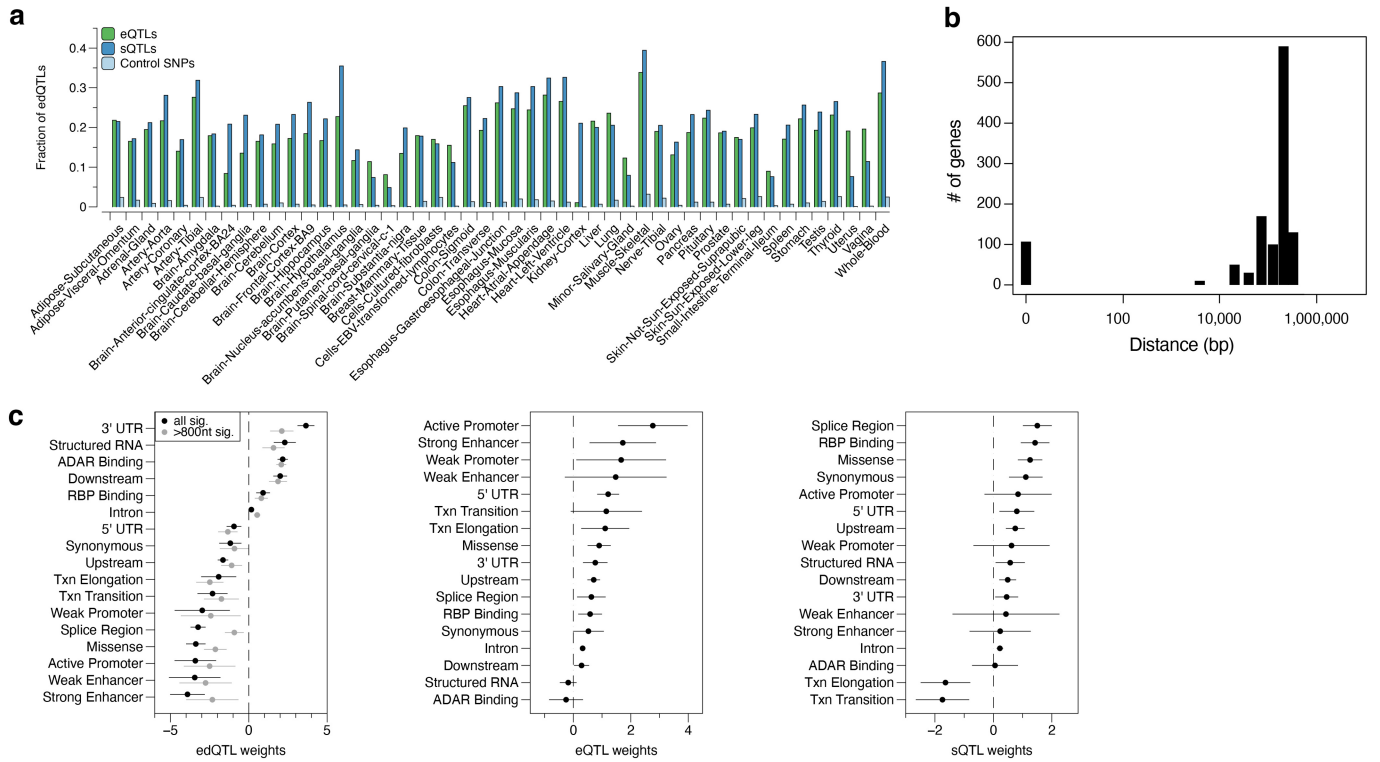


Extended Data Fig. 1 | Quantification of RNA editing levels in GTEx data. **a**, Number of editing sites used for *cis*-edQTL mapping in each tissue. Editing sites mapped uniquely by ≥20 reads in ≥60 samples of each tissue type were considered. Both GTEx V6p and V8 results are plotted here for comparison. **b**, Comparison of RNA editing levels quantified using exact read count versus estimated read count for *cis*-NAT editing sites (Methods). To maximize the number of tested *cis*-NAT editing sites, we collectively analyzed strand-specific RNA-seq data of 5 human tissues (spleen, lung, thyroid, colon and skin) by pooling the raw reads together. In total, $n = 806$ editing sites of 47 *cis*-NAT dsRNAs were used for comparison (≥20 reads coverage). **c**, Percentage of editing variance across individuals explained by top 10 principal components (PCs) of editing level. **d**, Association between *ADAR1* expression level with PC 1 of editing level in $n = 175$ Brain – cerebellar hemisphere samples. **e**, Density plot of editing level measurements and editing level variance between individuals ($n = 175$; normalized standard deviation) of 123,707 sites in Brain – cerebellar hemisphere tissue. Editing sites detected in ≥60 individuals were used for this analysis.

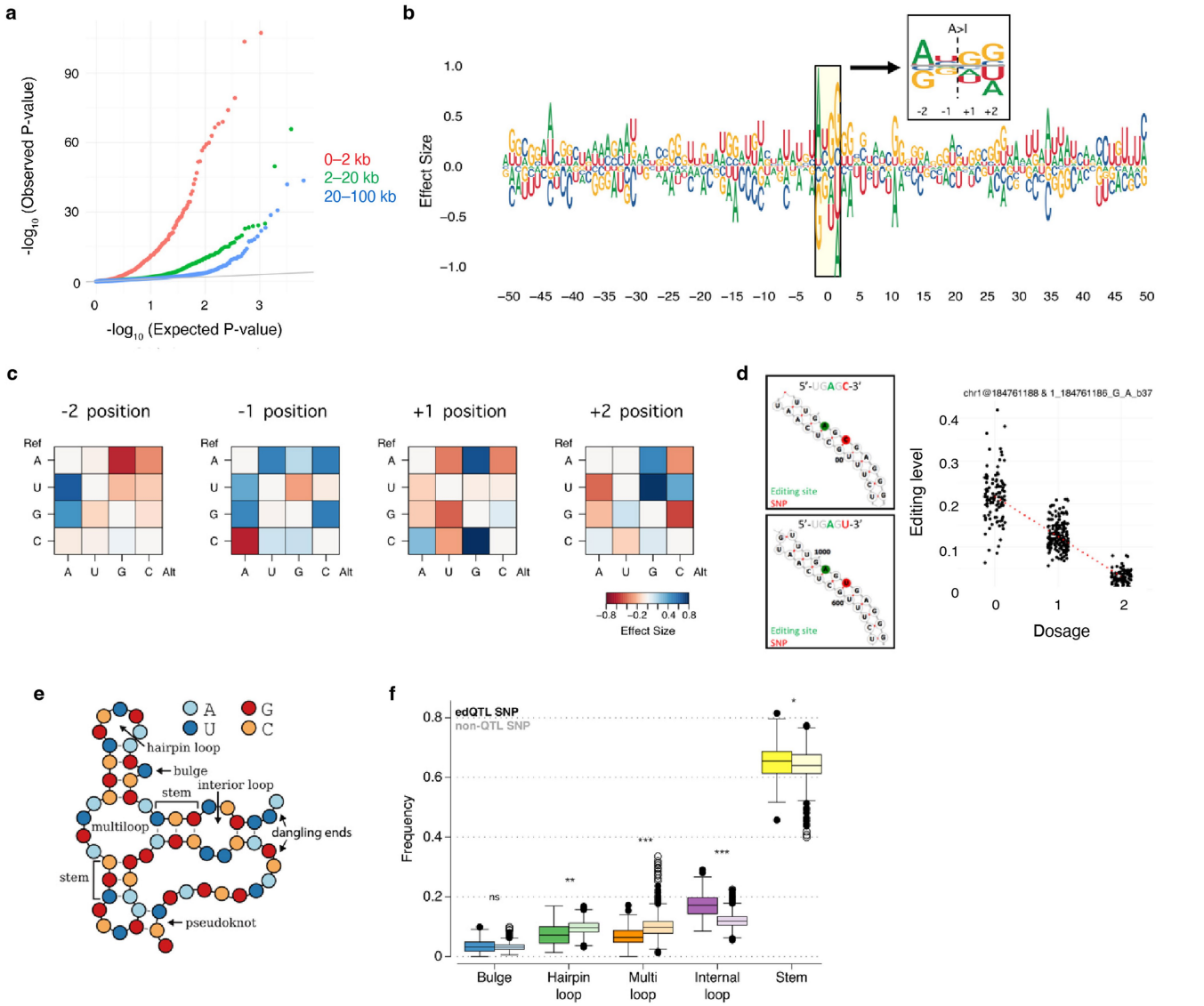


Extended Data Fig.2 | See next page for caption.

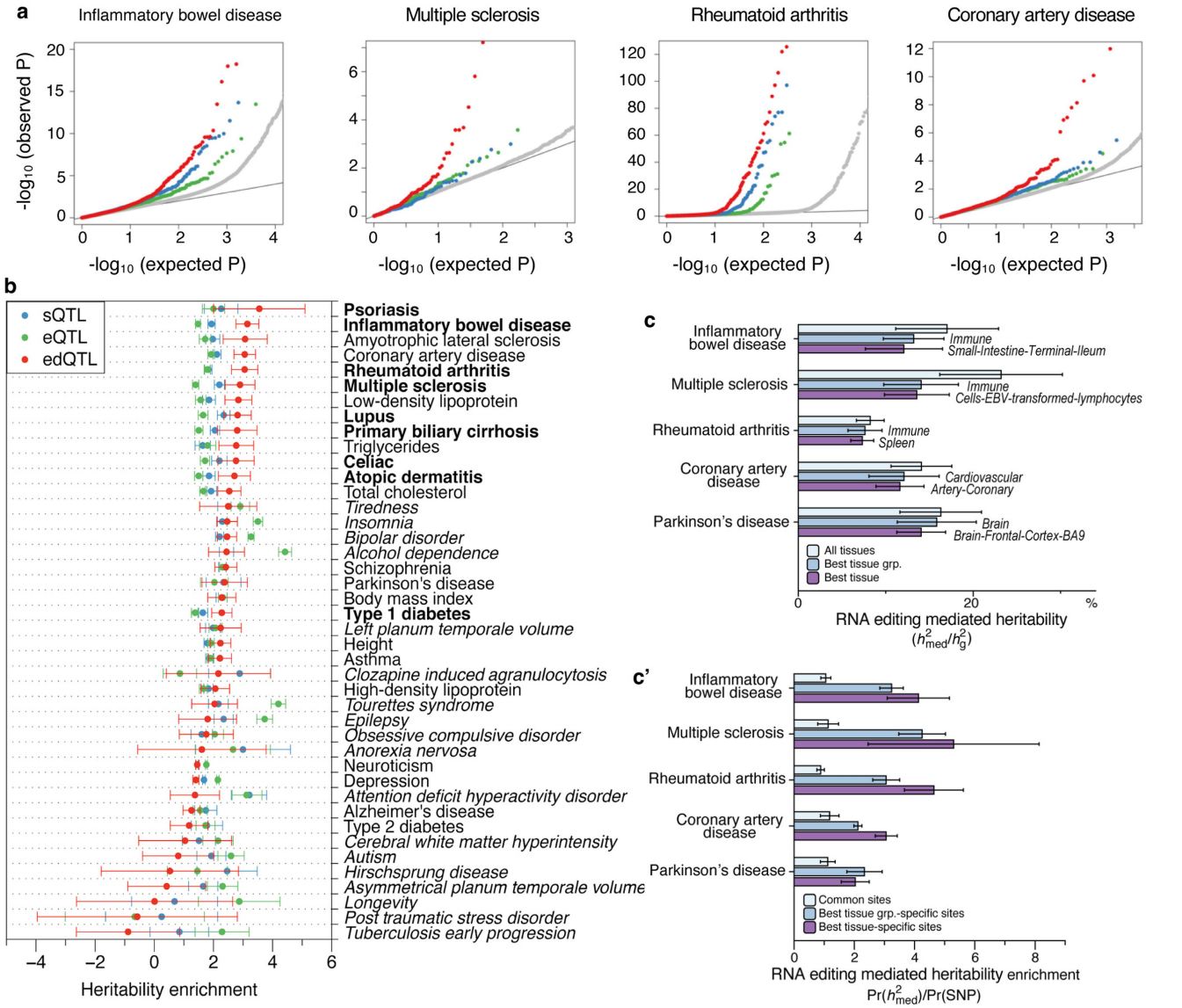
Extended Data Fig. 2 | Identification of RNA editing QTLs across GTEx tissues. a, Fraction of editing sites (left) and edited genes (right) found with edQTLs across *n* = 49 GTEx tissues. The LCLs data was obtained from previous edQTL study[33]. **b**, Fractions of edQTLs shared between multiple editing sites. **c**, Fractions of edGenes having multiple independent edQTLs. Each of the tissues (*n* = 49) was tested and plotted individually in **b** and **c**. **d**, Exemplary locus of *TRMT9B* showing two independent edQTLs each regulating multiple editing sites. The two edQTLs are represented by their lead SNPs rs13268982 and rs34995506, respectively. Genome browser view of the TRMT9B 3′ UTR that consists of four IRAlus (shown in the 1st track from top), with editing sites (2nd track from top), SNP location (black bars) and variant-editing association effect sizes for each editing site (3rd and 4th tracks) shown from top to bottom. Effect sizes are colored by direction of effects: red for positive effect and blue for negative effect. **e**, Comparison of the number of edSites identified in this study and in Park et al., 2021. For each tissue type, edSites are compared between two studies by evaluating (1) whether they are tested for genetic association, and (2) whether the associations are significant under different p-value cut-offs (p < 1e-3 on the left, p < 1e-5 on the right). **f**, Sharing of edQTLs across tissues by sign (same direction of effects). Euclidean distance matrix across tissues was calculated for hierarchical clustering using Ward's method. **g**, Fraction of edQTLs shared by the number of tissues according to sign. **g'**, Fraction of edQTLs shared by the number of tissues according to magnitude. Effects with more than 2-fold change of sizes from one another are considered as of different magnitudes. **h**, Number of tissue-specific edQTLs effects found in 20 representative tissues. Representative tissues were picked by two criteria: 1) large sample size (≥150); 2) least number of shared editing sites with other tissue. For example, a high number of editing sites are shared between three arteries tissues so we choose Artery-Coronary tissue as the representative one since it has the largest sample size.



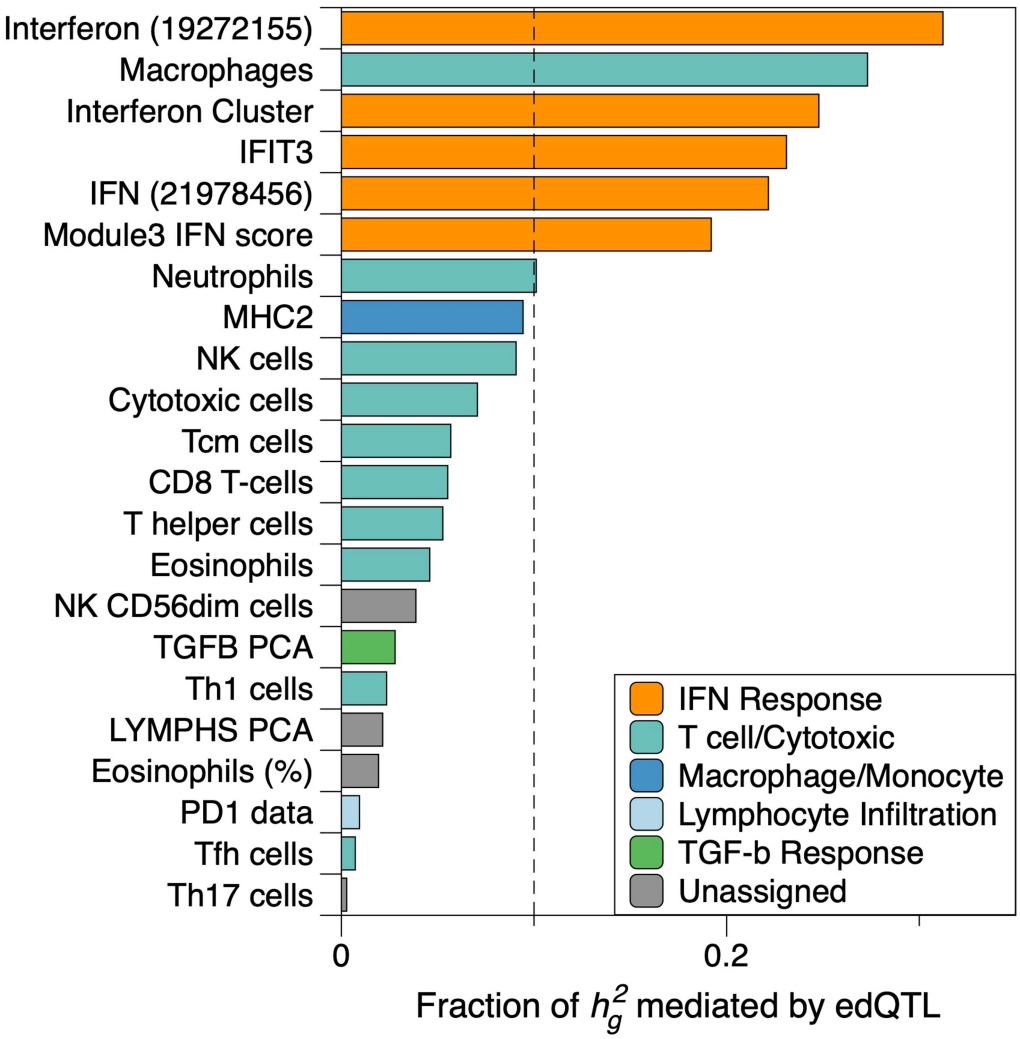
Extended Data Fig. 3 | Comparison between edQTLs, eQTLs and sQTLs in GTEx tissues. a, Sharing of edQTLs with eQTLs and sQTLs. We estimated the fraction of shared QTLs according to Storey's π_1 (Methods). SNPs with matching numbers and allele frequencies with the edQTLs were randomly sampled genome-wide in each tissue to be used as the control set. **b**, Distance between the best edQTL and best eQTL for genes with both types of QTL, using 1,478 genes in whole blood tissue as an example. **c**, Enrichment of functional elements underlying edQTLs, eQTLs and sQTLs (left to right). We used a Bayesian hierarchical model to identify putatively causal variants driving a QTL from a set of variants associated with the locus, and quantified the enrichment of strongly associated variants in functional elements. For edQTLs, we additionally identified the functional elements underlying edQTL SNPs that are >800 bp away from the edSites (grey dots). We used chromHMM annotations and gene-level annotations (e.g. intronic variant, splice region variants, UTR variants, etc.) from snpEff. Splice region variants include variants located within the region of splice site (1-3 bases of the exon side or 3-8 bases of the intron side) and branch point. Structured RNA are defined as regions with icSHAPE score ≥0.7 in RNA structure mapping data *in vivo* from multiple cell lines obtained from http://rasp.zhanglab.net/. ADAR1 and other RBP binding sites are defined using CLIP peaks obtain from http://postar.ncrnalab.org/. Error bars represent 95% confidence intervals. edQTLs n = 30,319; eQTLs n = 24,740; sQTLs n = 14,424.



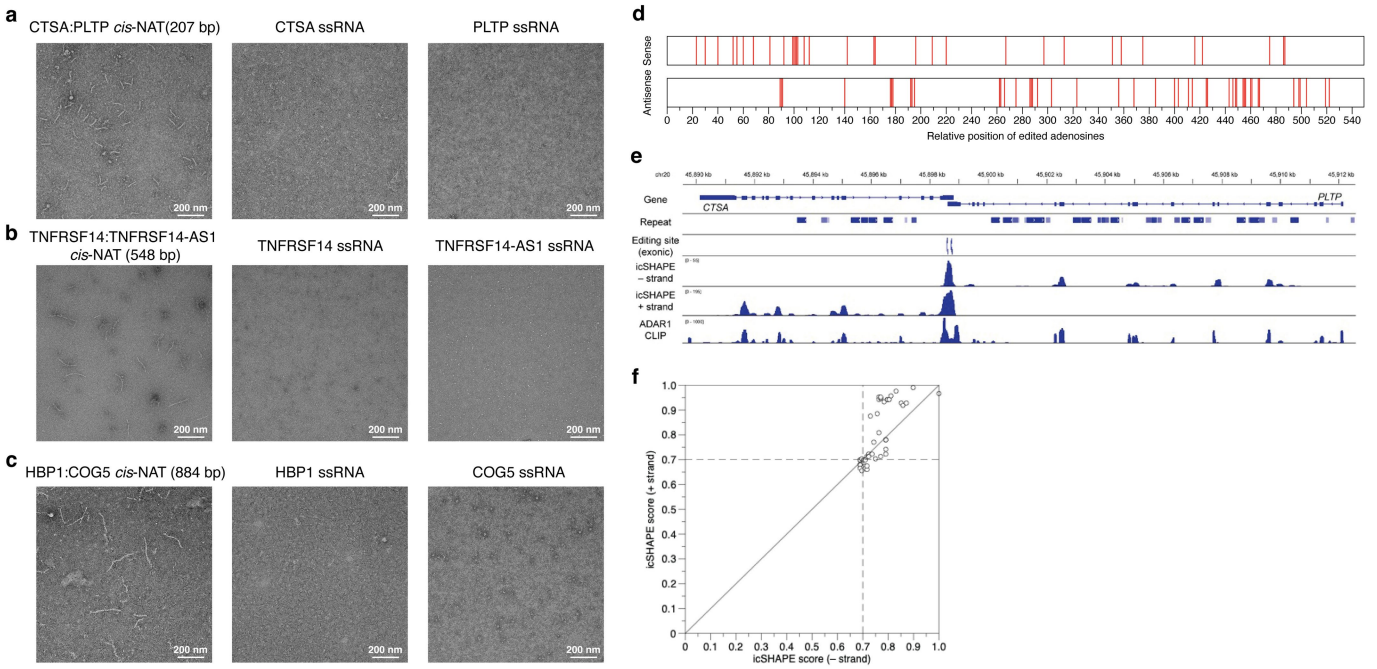
Extended Data Fig. 4 | Characterization of edQTLs' effects on RNA sequences and secondary structures. a, Q-Q plot of edQTLs (n = 30,319) annotated with distance from the associated editing sites. **b**, Summarized edQTL effect sizes of alternative alleles by 4 types of RNA ribonucleotides at each position from −50 to +50 nt relative to the editing site. Stronger effects were observed for ±2 and ±1 positions and illustrated to show the "AUAGG" motif centered at the edited "A". **c**, Further breakdown of the effects of 12 different nucleotide changes (reference allele -> alternative allele) caused by SNPs located at ±2 and ±1 positions. Reference alleles (y-axis) and alternative alleles (x-axis) were accounted for by their transcribed ribonucleotides on the RNA. **d**, An exemplary edSite association showing negative effects of C-to-U change at +2 position of editing site. Editing site at chr1: 184761188 and SNP at chr1: 184761186 are presented here (hg38 coordinates). Predicted RNA secondary structures for reference allele (G on DNA and C on RNA) and alternative allele (A on DNA and U on RNA) are shown on the left, with editing level measurements of different alternative allele dosages on the right. Biologically independent sample size n = 138. **e**, Schematic diagram showing RNA secondary structural features defined by bpRNA[66]. **f**, Frequency of edQTL SNPs found in different structural features (n = 142) compared to non-QTL SNPs (n = 4,080) in the same predicted structure. Statistical significance was calculated using two-sided Mann–Whitney U test. Box plots show interquartile ranges and median, with whiskers extending to minima and maxima.



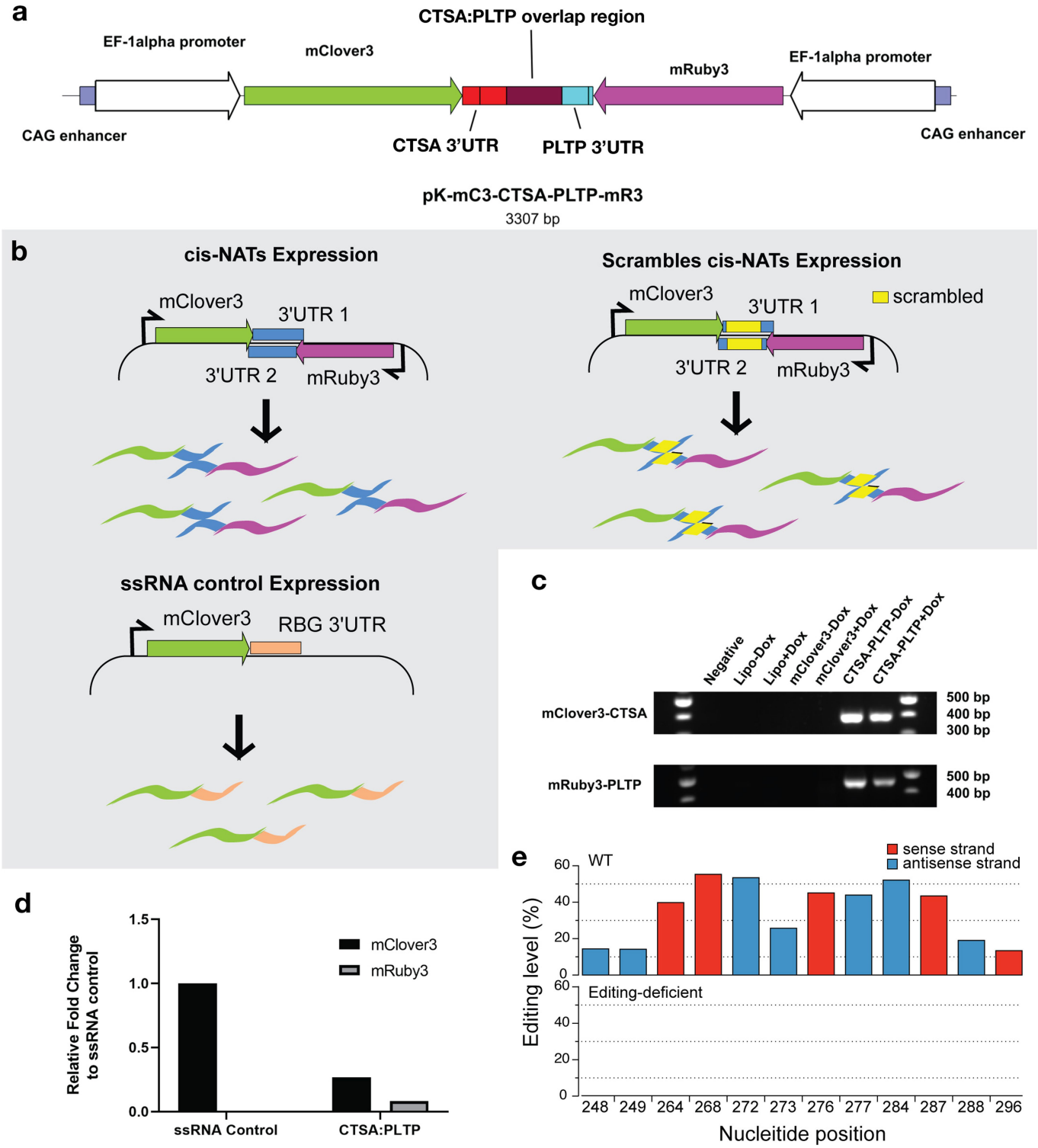
Extended Data Fig. 5 | RNA editing QTLs are enriched in immune-related diseases and immune traits. a, QQ-plots of IBD, MS, RA and CAD GWAS with QTL annotations. The expression levels of eGenes and sGenes were matched to the levels of edGenes within ±15% of deviation (see Methods). **b**, Enrichment of heritability for 42 human traits and diseases mediated by edQTLs, sQTLs and eQTLs. Autoimmune diseases are shown in bold. Enrichment is measured as the regression coefficient of MESC (Methods). Error bars represent standard deviation with mean as center. Meta-analysis of RNA editing QTL mediated heritability is shown in **c** and heritability enrichment in **c'**, for 4 exemplary autoimmune and immune-related diseases. Tissue groups are defined the same way as in Fig. 2c. Tissues and tissues groups shown in **c'** are the same as in **c**. Error bars represent standard deviation with mean as center. Sample sizes of studies are in Supplementary Table 1.



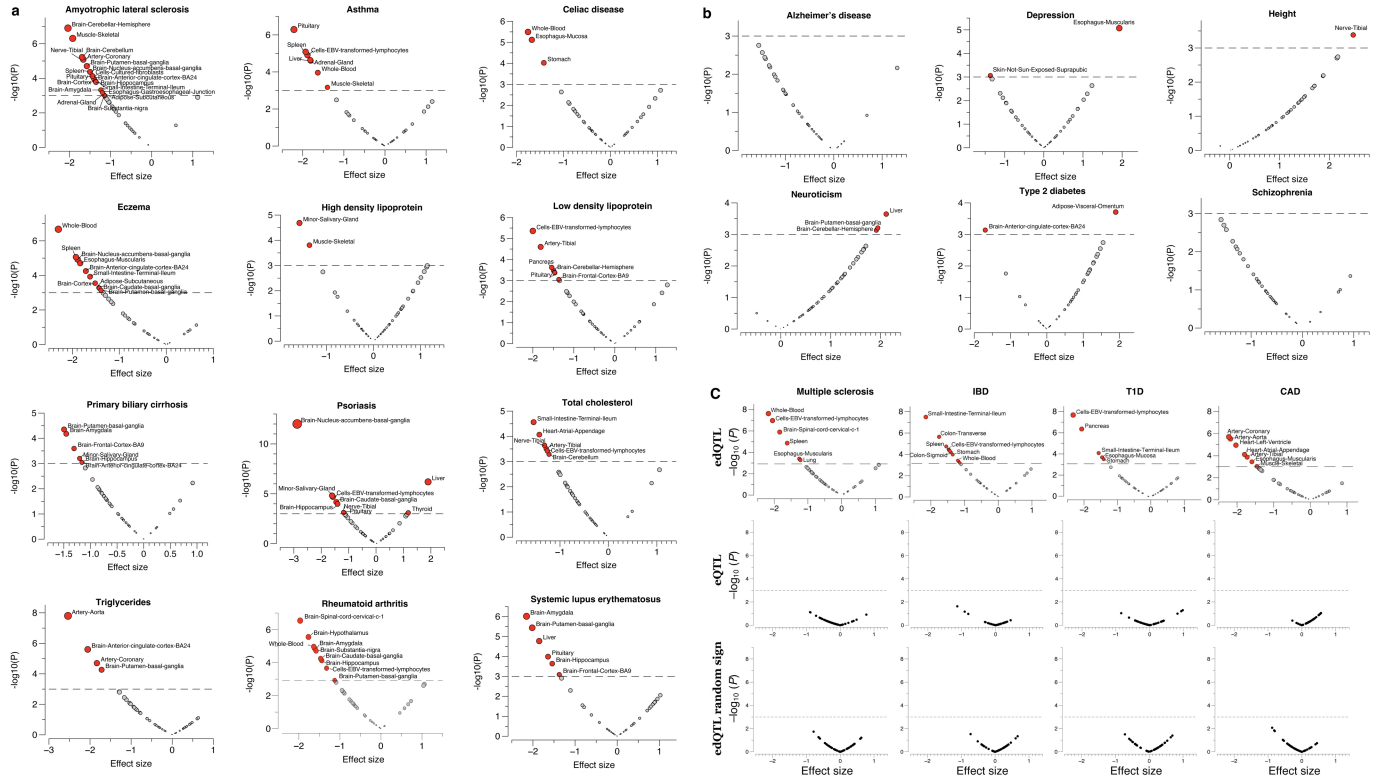
Extended Data Fig. 6 | Fraction of SNP heritability (h^2_g) mediated by edQTL in immune-related trait GWAS. We performed MESC analysis on 33 GWAS data (n = 9,138 individuals) obtained from Sayaman et al.[42]. in the same way as in Fig. 2C. We filtered out 11 immune traits with low total SNP heritability ($h^2_g < 0.01$) from the final result. The remaining 22 traits are categorized and colored by immune function annotations according to Sayaman et al.[42]. Dashed line indicates 0.1 of mediated SNP heritability (h^2_g)[42].



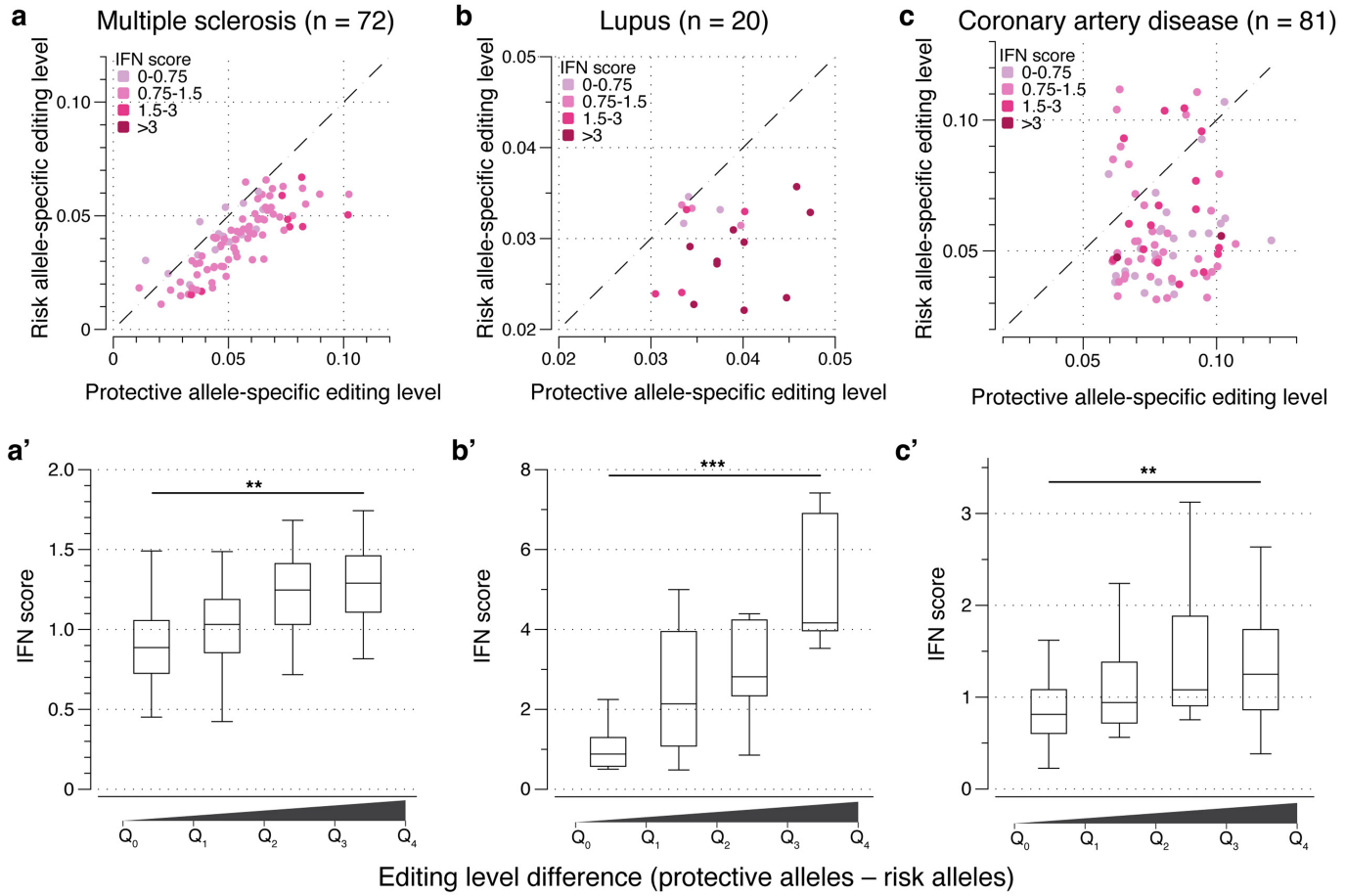
Extended Data Fig. 7 | Formation of *cis*-NAT dsRNAs *in vitro* with MDA5 proteins and in human cells. For (**a**) *TNFRSF14:TNFRSF14-AS1*, (**b**) *CTSA:PLTP* and (**c**) *HBP1:COG cis*-NAT dsRNAs, negative stain electron microscopy (EM) images are shown for MDA5 proteins incubated with *cis*-NAT dsRNAs (left), sense strand ssRNA control (middle) and antisense strand ssRNA control (right), respectively. All filamentation experiments were repeated independently for 3 times for each *cis*-NAT pair with similar results. **d**, *In vitro* editing status of dsRNAs formed by *TNFRSF14:TNFRSF14-AS1 cis*-NAT. Editing sites in both sense (*TNFRSF14*, top) and antisense (*TNFRSF14-AS1*, bottom) strands are shown as red bars separately for the two strands. Editing information was determined by Sanger sequencing. **e**, Genome browser snapshot of the *CTSA:PLTP cis*-NAT locus. Annotations of gene structure, *Alu* repeats, exonic editing sites, RNA structure mapping signal (icSHAPE scores in + and – strands) and ADAR1 CLIP signal are shown from top to bottom. **f**, Correlation of icSHAPE scores between sense and antisense strands for *cis*-NAT dsRNAs ($n = 38$). icSHAPE data was obtained from the RASP database: http://rasp.zhanglab.net/. ADAR1 CLIP data of human U87 cell lines was obtained from the POSTAR3 database: http://postar.ncrnalab.org/.



Extended DataFig. 8 | Overexpression of *CTSA:PLTP cis*-NAT in human cells. **a**, Schematic diagram showing plasmid construction for bidirectional transcription of *CTSA:PLTP cis*-NAT. Expression is driven by two opposing EF-1alpha promoters. Sense strand (*CTSA* 3'UTR) and antisense strand (*PLTP* 3'UTR) of the *cis*-NAT are downstream of mClover3 and mRuby3, respectively, with 207 bp overlapping sequences to mimic *cis*-NAT dsRNA formation. **b**, Cartoon showing vector design expressing *cis*-NAT, its scrambled formation and ssRNA control. **c**, RT-PCR data for validating the transcription of sense (*CTSA*) and antisense (*PLTP*) strands of *CTSA:PLTP cis*-NAT under different treatments. Amplicons (spanning mClover3-CTSA) are expected only when *CTSA* is expressed, and amplicons (spanning mRuby3-PLTP) are expected only when *PLTP* is expressed. Experiments were repeated independently for 2 times with similar results. **d**, Real-time PCR measurement of overexpression of *CTSA:PLTP cis*-NAT relative to mClover3 ssRNA control. Overexpression of sense (*CTSA*) and antisense (*PLTP*) strands was measured by the co-expressed mClover3 and mRuby3, respectively. n = 1 independent experiment. **e**, Editing level of *CTSA:PLTP cis*-NAT in HEK293 cells with WT ADAR1. Editing level was measured by Sanger sequencing.



Extended Data Fig. 9 | Directional effects of risk variants associated with complex traits and diseases on RNA editing levels. a, Traits and diseases with overall negative effects on RNA editing (increased risk associated with reduced editing level). **b**, Traits and diseases with overall positive or unclear direction of effects on RNA editing. **c**, Directional effect analysis of four exemplary diseases for edQTLs (top row), eQTLs (middle row) and control edQTLs with randomly flipped signs (bottom row). In all panels, we plotted the genome-wide −log10(P) against estimated effect size in GTEx tissues ($n = 49$), all estimated using SLDP regression. Significant tissues are colored in red (Bonferroni corrected multiple test p-value < $1x10^{-3}$). Larger circles denote lower p-values.



Extended Data Fig. 10 | Risk genetic variants are associated with reduced RNA editing levels and induced IFN score in inflammatory diseases. RNA-seq data from patient-derived samples of three diseases (multiple sclerosis, lupus and CAD) were used to calculate the overall editing levels of dsRNAs associated with risk vs protective alleles and the IFN scores. We used allele-specific editing analysis to determine editing levels associated with risk vs protective alleles (a, b and c). The IFN scores were calculated by summing up the expression of representative ISGs for each disease (Methods), and grouped in four different bins. Samples are then grouped by their editing level differences (protective alleles – risk alleles, divided by quartiles) to show IFN scores plotted for each group and compared between groups of the same disease type (a′, b′ and c′ with matching sample size as a, b and c, respectively). ***, $P < 0.001$; **, $P < 0.01$; one-way ANOVA test. Box plots in a′, b′ and c′ show interquartile ranges and median, with whiskers extending to minima and maxima.

nature portfolio

Corresponding author(s): Jin Billy Li

Last updated by author(s): 6/8/2022

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

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

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on 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 statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ 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

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

Software and code

Policy information about availability of computer code

Data collection	Codes used for editing level quantification in GTEx V8 data and the reference editing site list are available at: https://hub.docker.com/r/vanessa/mpileup/.
Data analysis	Codes availability: A full pipeline to preprocess the GWAS and edQTL data, prioritize relevant loci, run colocalization tests, and generate the associated plots is publicly available at https://github.com/mikegloudemans/rna-editing-coloc and https://github.com/vargasliqin/GTEx_edQTL. Third party softwares and codes: we used FastQTL (version v7, https://github.com/francois-a/fastqtl) to map cis-edQTLs. Tissue-wide comparison of edQTLs was performed using MashR v0.2 (https://github.com/stephenslab/mashr). STAR (v2.5.2b), BLASTN (v2.8.1) and trim_galore (v0.6.5) were used for ADAR1 CLIP-seq data analysis. ggseqlogo (v0.1) and bpRNA (https://github.com/padidehdanaee/bpRNA) were used for RNA motif and structure analysis. MESC (https://github.com/douglasyao/mesc) was used to estimate heritability mediated by edQTLs with the effect sizes estimated using LASSO (R package glmmLasso v1.5.1). SLDP (https://github.com/yakirr/sldp) was used for estimating directional effect of RNA editing in disease genetics. COLOC (v3.2-1) was used for colocalization analysis. RNA secondary structures were predicted using ViennaRNA (v2.4.14). SPRINT (v0.1.8) was used for hyper editing analysis. PLINK (v1.9) and SNPsnap (https://github.com/pascaltimshel/snpsnap/) were used to randomly select control sets of SNPs.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

GTEx v8 release editing level data and edQTL call sets are available on the GTEx Portal: https://www.gtexportal.org/home/datasets, under session: Single Tissue cis-RNA Editing QTL Data. Details of the GWAS summary statistics used for colocalization are provided in Supplementary Table 1. icSHAPE data was obtained from the RASP database: http://rasp.zhanglab.net/. CLIP data was obtained from the POSTAR3 database: http://postar.ncrnalab.org/.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-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. We required that at least 70 samples per tissue for edQTL mapping to ensure acceptable statistical power as indicated by data from https://www.science.org/doi/10.1126/science.aaz1776.
Data exclusions	We excluded donors without genotypes and tissues with < 70 samples from the edQTL analysis, which resulted in 49 tissue types and 838 donors. To ensure reliability of the editing level measurement, we only considered editing sites covered by ≥20 reads in ≥60 samples per tissue type for analysis. Details of the number of editing sites used for edQTL analysis are shown in Extended Data Fig. 1a.
Replication	We used a previous study of smaller sample size (doi:10.1186/s13059-017-1270-7) to validate our edQTL mapping result. We found 95.5% of the edQTLs are reproducible between our study and the other study with consistent direction of effects. For experimental data, we conducted at least 3 independent biological replicates and confirmed all attempts were successful.
Randomization	Donors are randomly allocated into tissue groups for edQTL analysis. To account for confounding effects of population, sex and age, we included the top 3 genotype PCs as well as sex and age as covariates for edQTL mapping.
Blinding	Since the analyses are based on genotypes within in each tissue group instead of individuals, investigators are naturally blinded.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T-ADAR1-E912A-iMDA5-mCherry-pIFN-Lucia cells were generated in the Li Lab at Stanford University from HEK293T cells sourced from ATCC

Authentication	None of the cell lines used were authenticated
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used