Global analysis of A-to-I RNA editing reveals association with common disease variants

Oscar Franzén[1], Raili Ermel[2,*], Katyayani Sukhavasi[3,*], Rajeev Jain[3], Anamika Jain[3], Christer Betsholtz[1,4], Chiara Giannarelli[5,6], Jason C. Kovacic[5], Arno Ruusalepp[2,3,7], Josefin Skogsberg[8], Ke Hao[6], Eric E. Schadt[6,7] and Johan L.M. Björkegren[1,3,6,7]

[1] Integrated Cardio Metabolic Centre, Karolinska Institutet, Huddinge, Sweden
[2] Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia
[3] Department of Pathophysiology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia
[4] Department of Immunology, Genetics and Pathology, Uppsala Universitet, Uppsala, Sweden
[5] Cardiovascular Research Center, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America
[6] Institute of Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America
[7] Clinical Gene Networks AB, Stockholm, Sweden
[8] Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna, Sweden
[*] These authors contributed equally to this work.

ABSTRACT

RNA editing modifies transcripts and may alter their regulation or function. In humans, the most common modification is adenosine to inosine (A-to-I). We examined the global characteristics of RNA editing in 4,301 human tissue samples. More than 1.6 million A-to-I edits were identified in 62% of all protein-coding transcripts. mRNA recoding was extremely rare; only 11 novel recoding sites were uncovered. Thirty single nucleotide polymorphisms from genome-wide association studies were associated with RNA editing; one that influences type 2 diabetes (rs2028299) was associated with editing in *ARPIN*. Twenty-five genes, including *LRP11* and *PLIN5*, had editing sites that were associated with plasma lipid levels. Our findings provide new insights into the genetic regulation of RNA editing and establish a rich catalogue for further exploration of this process.

INTRODUCTION

Originally discovered in trypanosomes (*Benne et al., 1986*), RNA editing is a co-/post-transcriptional process that increases RNA diversity in eukaryotes. The most common modification in mammals is the deamination of adenosine to inosine (A-to-I), catalyzed by adenosine deaminases acting on RNA (ADAR) enzymes (*Farajollahi & Maas, 2010*). Inosine base pairs with cytosine and is interpreted as guanosine by the cellular machinery

(*Bass, 2002*; *Farajollahi & Maas, 2010*). A-to-I editing occurs in many human tissues and acts on the majority of human pre-mRNAs (*Park et al., 2012*; *Bahn et al., 2012*; *Bazak et al., 2014*). Most RNA editing takes place in *Alu* repetitive elements. These short interspersed nuclear elements are ∼300 base pairs (bp) in length, specific to primates, and often embedded in introns and 3′ untranslated regions (UTRs). The abundance of *Alus* in the human genome—more than one million copies—makes it very likely to encounter inverted pairs of *Alus*, which can form stem-loop structures that are favored substrates of ADAR. Cytosine-to-uracil (C-to-U) editing is much less frequent than A-to-I editing (*Blanc & Davidson, 2003*) and is mediated by APOBEC enzymes. The medical implications of RNA editing may be extensive as it has been implicated in diseases ranging from cancer and neurological diseases (*Hideyama et al., 2010*; *Hood & Emeson, 2011*; *Galeano et al., 2012*; *Slotkin & Nishikura, 2013*) to atherogenesis (*Stellos et al., 2016*).

Detection of RNA-edited sites is relatively straightforward since the sequencing of an RNA-edited transcript (actually cDNA) will capture the edited base (*Ramaswami & Li, 2016*). A discrepancy in the alignment between the sequence and the genomic reference can be an RNA editing event, a single nucleotide polymorphism (SNP), or an artifact related to sequencing or data processing. Several computational methods have been devised to call RNA editing events from RNA-seq data (*Ramaswami et al., 2013*; *Picardi et al., 2015*; *Zhang & Xiao, 2015*; *Kim, Hur & Kim, 2016*; *Ramaswami & Li, 2016*; *Tan et al., 2017*).

Here, we analyzed transcriptome-wide RNA editing in one large cohort of patients with coronary artery disease—the S̲tockholm-T̲artu A̲therosclerosis R̲everse N̲etwork E̲ngineering T̲ask (STARNET) (*Franzén et al., 2016*) study. We implemented a novel detection pipeline with several systematic filtering steps to deliver a refined set of RNA editing events. Our pipeline borrows ideas from published methods and it is designed for being executed on a computer cluster.

First, we focused on patterns of global RNA editing. To identify clinical features associated with RNA editing, we mapped RNA editing quantitative trait loci (edQTLs). Finally, to assess the disease relevance of RNA editing, we analyzed our edQTLs in the context of published genome-wide association studies (GWAS).

METHODS

Statistics

R v.3.2.2 was used for all statistical analyses (*R Core Team, 2015*). If not otherwise specified, P-values were calculated with Welch's t-test. All statistical tests were two-sided.

Data sources

RNA-seq and genotype data (generated with the Illumina Infinium OmniExpressExome-8 chip) in this study originated from samples collected of up to 7 tissues (artery aorta, internal mammary artery, whole blood, subcutaneous fat, visceral fat, liver, and skeletal muscle) and 2 cell types (macrophages and foam cells) (*Franzén et al., 2016*) from 855 human subjects. Not all tissues/cell types were collected from all subjects. The rationale for collecting these tissues and their role in coronary artery disease have previously been described (*Björkegren et al., 2015*). Ethical permits, biopsy and experimental procedures

have previously been described (*Franzén et al., 2016*) (Karolinska Institutet Dnr 154/7 and 188/M-12). The samples from aortic wall ($n = 538$) and internal mammary artery ($n = 552$) were exclusively prepared from rRNA-depleted libraries generated with the Illumina RiboZero protocol. The remaining samples were almost all poly(A)-enriched. Of the 4301 samples included in this study, 52% (2,267/4,301) were sequenced with a strand-specific protocol, and the remaining were sequenced with a non-strand-specific protocol. A complete summary of tissues, samples and protocols is shown in Tables S1 and S2. Editing calls are listed in Dataset S1.

Metabolite and protein data were generated by Olink AB (Sweden) and Nightingale (formerly Brainshake, Finland).

Genome version, annotations and scripts

The human genome GRCh38 and GENCODE (*Harrow et al., 2012*) v.24 annotations were used throughout this study. RNA editing sites were mapped to genomic features with ANNOVAR (*Wang, Li & Hakonarson, 2010*) v.2015-03-22. miRBase (*Kozomara & Griffiths-Jones, 2014*) v.21 was used for miRNA annotations. Scripts used to process data can be retrieved from: https://github.com/oscar-franzen/rnaed.

Computational scheme for detection of RNA editing sites

Sequencing reads were aligned with the human genome using STAR (*Dobin et al., 2013*) v.2.5.1b, with the maximum number of alignments set to 1, and the ratio of mismatches to sequence length set to 0.1. Only uniquely mapped reads were considered (mapping quality = 255). The strand specificity of sequencing was confirmed by counting reads mapped to sense and antisense transcripts with HTSeq (*Anders, Pyl & Huber, 2014*) v.0.6.0. For all samples, the predicted strandedness matched the recorded information from the sequencing facility. Only samples with more than 1,000,000 uniquely mapped reads in genes were analyzed. The first step in our pipeline was collapsing polymerase chain reaction (PCR) duplicates with the rmdup command in samtools v.1.2. This step limits the number of false positives from the library construction process (i.e., amplification of errors introduced by the reverse transcriptase). Alignments were then scanned for single nucleotide variants (SNVs) by parsing the output of the mpileup command in samtools. SNVs were removed if located within simple repeats, homopolymers ≥ 5 bp, within 5 bp of splice junctions. SNVs located on the mitochondrial genome, on unplaced contigs, and within the major histocompatibility complex were also removed. Intersections were called with bedtools (*Quinlan & Hall, 2010*) v.2.21.0. Furthermore, SNVs listed in any of the following databases were discarded: NCBI dbSNP b141, b146, and b147; Exome Aggregation Consortium variants v.0.3.1; NHLBI ESP6500 variants; Scripps' Wellderly (*Erikson et al., 2016*); and COSMIC (*Forbes et al., 2015*) v.77. For the remaining SNVs, all sequencing reads overlapping the site were extracted and realigned with GSNAP (*Wu & Nacu, 2010*) v.2016-05-25, using settings that allow mismatches and splice junctions. For each sequencing read, the alignment coordinates (chromosome, strand, start and stop positions) from STAR and GSNAP were compared, and discordant alignments were removed. A concordant alignment was defined as >99% overlap of the start-stop interval

on the same chromosome and strand. To call an RNA editing event, the base quality score had to be ≥ 20. To avoid errors introduced from random hexamer priming, the SNV had to be >6 bases from the 5′ start position of the sequencing read. SNVs indicating more than one type of variant per site were ignored. The final criteria required at least two non-identical sequencing reads to support the RNA editing event.

121.2 billion single-end reads were mapped to unique locations in the human genome (median = 27,539,768 reads/sample). Collapsing PCR duplicates reduced the numbers to ∼42.9 billion aligned reads (median = 9,527,872 reads/sample). In total, 22,124,713 RNA editing events were called, mapping to 2,631,392 unique sites. 83.3% (2,866,471/3,438,790) and 92.6% (17,304,283/18,685,923) of the identified events were A-to-G mismatches in strand-specific and non-strand-specific libraries, respectively. In the non-strand-specific data, these numbers also include the reverse complement of A-to-G (i.e., T-to-C).

Quality of detected editing events

Sequencing errors are more likely to occur toward the end of reads. We did not find 5′ or 3′ positional biases in sequencing reads for canonical events (A-to-I and C-to-U; Fig. S13). Next, assuming that G-to-A events in strand-specific samples were sequencing or mapping errors, we estimated that the false discovery rate (FDR) was 3.1% (89,490/2,866,471). For non-strand-specific samples (assuming all G-to-C events were sequencing or mapping errors), the FDR was 0.83% (72,500/8,717,781).

Collected RNA samples from macrophages and foam cells are near biological replicates (foam cells are derived from macrophages by incubating them with acetylated LDL for 48 h). Therefore, we assessed the reproducibility of RNA editing in 235 macrophage—foam cell pairs. We first assessed the percentage of editing sites detected in both samples in each pair; the median across all pairs was 24%. It may therefore be plausible that many RNA editing sites are not strictly reproducible even in samples that are close to biological replicates. This may reflect the stochastic nature of RNA editing or variability in sequencing depth. We also determined whether the editing ratios were stable for the ∼24% of sites found in both samples of each pair. Across each of the 235 macrophage-foam cell pairs, the median Spearman correlation was 0.83 (Fig. S14). Thus, despite known technical and biological confounders, the reproducibility was relatively high for sites that can be detected in both samples.

Gene expression

Gene expression was calculated as RPKM (*Mortazavi et al., 2008*) from reads counted with HTSeq (*Anders, Pyl & Huber, 2014*).

Association-testing of editing ratios and clinical parameters

Association-testing was conducted with the linear model in MatrixEQTL (*Shabalin, 2012*) v.2.1.1. Clinical parameters were first rank-normal transformed with the function *rntransform* in the R package GenABEL (*Aulchenko et al., 2007*) with ties broken randomly. Two biological covariates (age and sex) and two technical covariates (sequencing batch and laboratory) were included in the statistical model. Only sites with non-zero A-to-G(I) editing ratios in more than 50 samples per tissue were considered. Significant associations

were defined as those satisfying the following criteria: (i) adjusted $P < 0.05$ (10,000 permutations were executed for each clinical parameter-tissue combination); (ii) R-sq > 0.4; and (iii) no correlation between gene expression (of the overlapping gene) and the tested clinical parameter (nominal $P > 0.05$).

RNA-editing quantitative trait loci (edQTLs)

Genotypic data were processed as described (*Franzén et al., 2016*); biallelic markers (MAF > 0.05) were converted to allele dosages (0, 1, and 2). edQTLs were generated separately for each tissue. Only A-to-G(I) sites found (i.e., editing ratio > 0) in more than 40 samples were included in the analysis (sites on sex chromosomes were excluded). Site-sample combinations with fewer than <5 reads supporting editing were removed. Editing ratios were then rank-normal transformed as described, and the null hypothesis of no association between allele dosages and editing ratios was tested with a linear regression model in MatrixEQTL (*Shabalin, 2012*). For each hypothesis, we required at least 20 subjects in at least 2 genotype-allele groups. Thus, the minimum number of individuals for any test was 40. All possible genotype-editing site combinations were tested regardless of the absolute genetic distance. We corrected for the following covariates: age, sex, laboratory, batch, and population stratification (i.e., cryptic relatedness of study subjects; represented as four genetic multi-dimensional scaling components). The FDR was controlled using a permutation procedure (*Franzén et al., 2016*): the RNA editing matrix was scrambled 1,000 times, i.e., breaking the link between sample identifiers and corresponding measurements. For each editing site, we generated 1,000 permutation *p*-values, which were used to adjust the real *p*-values. An adjusted *p*-value <0.20 was considered significant.

Overlap with eQTLs

Normalized gene expression data from (*Franzén et al., 2016*) were used in a linear regression model. *P*-values were adjusted with the Benjamini–Hochberg procedure and adjusted $P < 0.05$ was considered significant. For each editing site, all overlapping genes were tested.

GWAS SNPs

The NHGRI-EBI GWAS catalog (*Welter et al., 2014*) (r2018-01-01) and GWASdb (*Li et al., 2016*) v.2 were downloaded and merged into one database, keeping only SNPs with $P < 1e-08$. GWAS SNPs that were also edSNPs were saved in a new list that was LD-pruned with PLINK (–indep-pairwise 400 kb 1 0.6) (*Purcell et al., 2007*).

RESULTS

General characteristics of the RNA editome

To characterize the human RNA "editome", we analyzed RNA-seq data from the STARNET (*Franzén et al., 2016*) study, in which samples were obtained from up to 9 tissues (atherosclerotic aortic wall, $n = 538$; non-atherosclerotic arterial wall (internal mammary artery), $n = 552$; liver, $n = 545$; skeletal muscle, $n = 533$; visceral abdominal fat, $n = 533$; subcutaneous fat, $n = 532$; whole blood, $n = 559$; macrophages, $n = 256$; and foam cells, $n = 235$). In addition, 18 coronary artery samples were included. Sequencing was
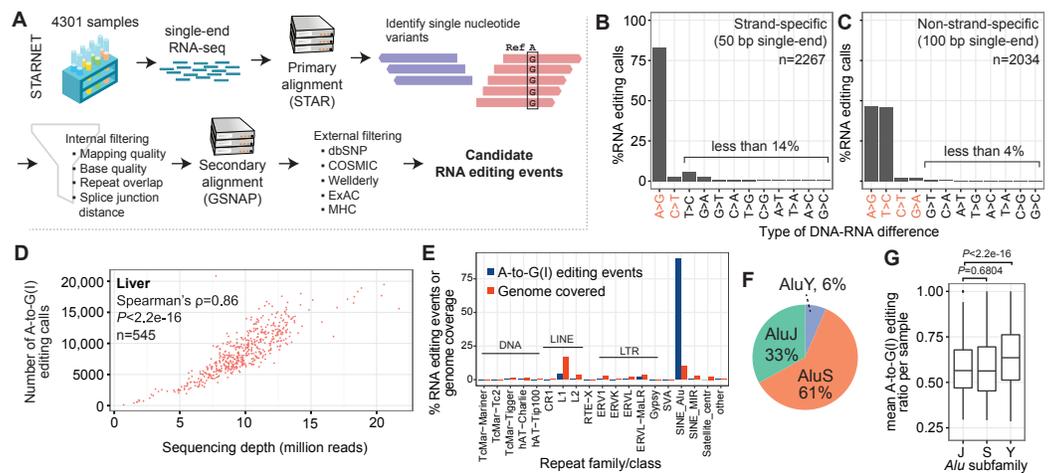
Figure 1 **General characteristics of RNA editing events.** (A) Simplified flowchart of the data analysis pipeline. See Fig. S1 for details. MHC, major histocompatibility complex. (B) Percentages of RNA editing types that were identified. The *x*-axis shows the editing type, and the *y*-axis shows the percentage of total editing calls. (The plot counts redundant sites, i.e., if the same RNA editing site is found in two samples, then the site is counted twice.) Editing events were identified in strand-specific libraries. Changes indicated in red (*x*-axis) are canonical events. The majority of events are consistent with A-to-I editing. (C) As (B) except that editing events were identified in non-strand-specific libraries. (D) Relationship between sequencing depth after collapsing of PCR duplicates (*x*-axis; million reads) and number of A-to-G(I) editing calls (*y*-axis). Each dot represents a liver sample from one subject. Only the non-strand-specific samples are shown. (E) Percentage of A-to-G(I) events in various genomic repeats. The repeat family/class is shown on the *x*-axis and identified events/genomic coverage in percent is on the *y*-axis. *Alu* is the only repeat class that shows enrichment compared with the total genomic coverage. (F) Pie chart of A-to-G(I) events in *Alu* subfamilies (S, Y, and J). (G) The box plot shows sample mean editing ratios of A-to-G(I) events. One outlier was removed from the Y subfamily. Editing events in *Alu* Y have significantly higher editing ratios compared with those in *Alu* J and S (Mann–Whitney *U* test; $P < 2.2e-16$).

Full-size ☒ DOI: 10.7717/peerj.4466/fig-1

done by using 50-bp single-end reads (these libraries were strand-specific; $n = 2,267$) and 100-bp single-end reads (these libraries were non-strand-specific; $n = 2,034$). Differences in read-length and strand specificity affected the sensitivity to detect RNA editing, and most analyses were therefore done separately on these two sets of samples. See Table S1 for a summary of analyzed samples. We applied a rigorous computational pipeline that includes internal and external filters to identify RNA editing in a total of 4,301 RNA-seq samples (Fig. 1A; Fig. S1). The number of samples per tissue ranged from 235 in foam cells to 559 in whole blood. For each putative editing site, we calculated the RNA editing ratio, defined as the count of edited bases divided by the total number of bases covering the site. The RNA editing ratio takes values within the semi-closed interval (0,1].

Most of the identified events were A-to-G mismatches (Figs. 1B and 1C; Table 1; Tables S2 and S3), consistent with A-to-I editing; henceforth we refer to these events as A-to-G(I) editing (including the reverse complement when the strand information was not preserved). We identified 1,484,357 and 450,002 A-to-G(I) events in non-strand-specific and strand-specific libraries, respectively; the number of non-redundant A-to-G(I) editing events in non-strand-specific and strand-specific libraries was 1,688,815. More A-to-G(I)

Table 1 Summary of data and RNA editing events.

Library type	Seq. read length (bp)	No. samples	Seq. depth ($\times 10^9$)[a]	Total no. RNA editing events	No. A-to-G(I) events	No. C-to-T(U) events
Strand-specific	50	2,267	20.6	714,372	450,002	40,116
Non-strand-specific	100	2,034	22.3	2,217,412	1,484,357[b]	375,935[c]

Notes.

[a]Total read count after collapsing of PCR duplicates (billion single-end sequencing reads).

[b]Includes T-to-C calls.

[c]Includes G-to-A calls.

events were detected in non-strand-specific libraries, owing to deeper sequencing and longer read length. Liver had the highest number of A-to-G(I) events, possibly reflecting higher transcriptional complexity. In total, 53.9% (910,943/1,688,815) of identified A-to-G(I) events had previously been reported in REDIportal (*Picardi et al., 2017*) or DARNED (*Kiran & Baranov, 2010*) (Fig. S2). Aortic wall had the highest number of novel A-to-G(I) sites ($n = 117,493$). A-to-G(I) editing was detected in transcripts from 25,951 genes, of which 62.1% (16,131/25,951) were protein-coding. We intersected RNA editing events with genome annotations (Fig. S3), showing that 56.0% of A-to-G(I) events were in introns, 18.3% in intergenic regions, and 9.4% in 3′ UTRs; 67.7% of the A-to-G(I) editing sites were tissue-specific when comparing exact chromosome-positions without considering differences in gene expression (Fig. S4). To exclude bias from differences in gene expression we repeated the analysis and only examined genes being robustly expressed in 7 tissues (defined as median reads per kilobase of transcript per million mapped reads [RPKM] >10 across all samples; 2,639 genes), the result showed that most events were tissue-specific (Fig. S4).

In strand-specific samples, the second most common change was noncanonical T-to-C editing (16.2%, 116,318/714,372). Interestingly, in strand-specific libraries the number of A-to-G(I) events correlated with the number of T-to-C events (Spearman's $\rho = 0.67$, $P < 2.2e{-}16$). T-to-C events have been reported in studies utilizing high-throughput sequencing data but it is unknown whether T-to-C events are authentic or artifacts (*Bahn et al., 2012*).

The relationship between sequencing depth and number of editing events was nearly linear (Fig. 1D and Fig. S5; liver samples: Spearman's $\rho = 0.86$, $P < 2.2e{-}16$), suggesting that detection of RNA editing per sample was not saturated or that the mechanisms mediating RNA editing are nonspecific. If RNA editing is nonspecific, it can be assumed that deeper sequencing does not saturate detection. Analysis of the cumulative number of unique A-to-G(I) sites as a function of added samples did not reveal a clear tendency toward saturation (Fig. S6A). It is also possible that saturation cannot be detected because most editing takes place in introns, which are much less captured with RNA-seq. However, analysis of the number of genes in which RNA editing occurred did reveal a trend toward saturation (Fig. S6B).

Differential editing in *Alu* subfamilies

In the present version of the human genome assembly (GRCh38) there are 1,186,920 *Alus* covering 10.2% (315,412,481 bp/3,088,269,832 bp). *Alu* was the sole element significantly enriched in editing events (Fig. 1E; Figs. S7–S9). Overall, 70.3% (1,187,358/1,688,815) of all A-to-G(I) events were in *Alu*s. Of the *Alus* in the human genome, 20.3% (241,226/1,186,920) had at least one A-to-G(I) editing event. Despite covering 20.7% (641,953,033 bp/3,088,269,832 bp) of the genome (*Lander et al., 2001*), LINE elements contained only 4.0% (67,810/1,688,815) of the A-to-G(I) events. The mean editing ratio per sample was slightly lower in non-*Alu* A-to-G(I) compared to editing sites in *Alus* ($\text{mean}_{\text{non-Alu}} = 0.54$, $\text{mean}_{\text{Alu}} = 0.57$, $P = 3.5e-15$), likely reflecting the strong preference of ADAR for double-stranded folds. When *Alu* elements were grouped by subfamilies (Y, J, and S) according to their established phylogeny (*Bennett et al., 2008*), it showed that 42.5% (718,653/1,688,815) of A-to-G(I) editing events were located within the *Alu*S subfamily, followed by *Alu*J (23.2%) and *Alu*Y (4.5%), see Fig. 1F. The mean editing ratio per sample was highest in *Alu*Y ($\text{mean}_{AluS} = 0.56$, $\text{mean}_{AluJ} = 0.56$, $\text{mean}_{AluY} = 0.64$, $P < 2.2e-16$; Fig. 1G). Mean editing levels were also significantly higher within nongenic *Alus* compared to genic *Alus* ($\text{mean}_{\text{genic Alus}} = 0.56$, $\text{mean}_{\text{nongenic Alus}} = 0.66$, $P < 2.2e-16$), perhaps because purifying selection lowered editing levels in mRNAs.

ADAR1 is ubiquitously expressed and *ADAR2* displays a vascular-specific signature

ADAR1 and ADAR2 catalyze A-to-I editing (*Macbeth et al., 2005*; *Nishikura, 2016*). ADAR3 is catalytically inactive and may have regulatory properties (*Chen et al., 2000*). The expression of *ADAR1*, *ADAR2*, and *ADAR3* was compared across the tissues/cell types (Fig. 2; Fig. S10). *ADAR1* was uniformly expressed in all tissues (median = 31 RPKM, median absolute deviation [MAD] = 14); the lowest levels observed were in skeletal muscle (median = 5 RPKM, MAD =1), see Fig. 2A. The latter was also observed in GTEx (*Melé et al., 2015*). Expression of *ADAR1* correlated with A-to-G(I) editing events in every tissue except foam cells (q-value < 0.05), further corroborating the authenticity of called RNA editing events. Whole blood showed the strongest correlation between A-to-G(I) editing events and *ADAR1* expression (strand-specific samples, Spearman's $\rho = 0.85$, $P < 2.2e-16$; Fig. S11). Although sex-specific differences in *ADAR1* expression have been reported in tumor samples (*Paz-Yaacov et al., 2015*), we found only modest sex-specific differences in foam cells ($P = 0.015$), liver ($P = 0.049$), and subcutaneous fat ($P = 0.017$; Fig. S12).

ADAR2 expression was low in all tissues (Fig. 2B; median = 3 RPKM, MAD = 2) except the three vascular tissues (atherosclerotic aortic wall, internal mammary artery, and coronary artery; median = 23 RPKM, MAD = 14); expression was highest in internal mammary artery (median = 42 RPKM, MAD = 12). *ADAR2* expression correlated with the number of A-to-G(I) events only in atherosclerotic aortic wall (Spearman's $\rho = 0.18$, $P = 2.6e-05$) and was stronger in females (females: $\rho = 0.19$, $P = 0.012$; males: $\rho = 0.17$, $P = 5.2e-04$). However, there were no sex-specific differences in editing in any tissues ($P > 0.05$). *ADAR3* was essentially below the detection level in all tissues (median = 0.01 RPKM, MAD = 0.02).
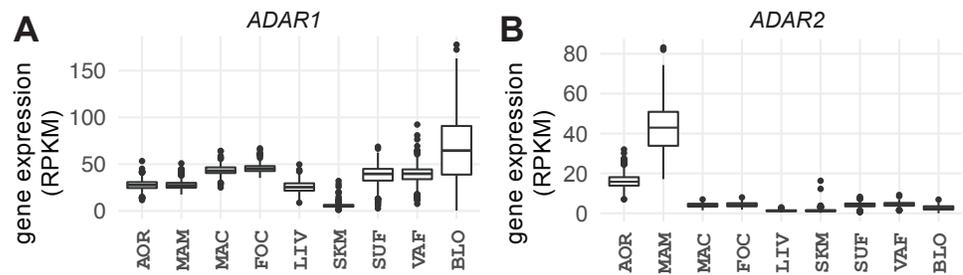
Figure 2 Expression of *ADAR1* and *ADAR2* across studied tissues. Box plots show gene expression of (A) *ADAR1* and (B) *ADAR2* in RPKM (*y*-axis). AOR, aorta; MAM, internal mammary artery; MAC, macrophage; FOC, foam cell; LIV, liver; SKM, skeletal muscle; SUF, subcutaneous fat; VAF, visceral fat; BLO, whole blood.

Full-size 🖼 DOI: 10.7717/peerj.4466/fig-2

Eleven novel recoding sites and one loss of stop codon

RNA editing provides a mechanism for cells to generate additional protein diversity (*Nishikura, 2010*). However, novel protein recoding events—those that cause amino acid changes—are rare. We systematically searched for A-to-G(I) editing events causing recoding. We found that 3.9% (67,429/1,688,815) of A-to-G(I) events were within known protein-coding exons (variants arising from antisense transcription were excluded), and only 3.0% (50,883/1,688,815) were non-synonymous substitutions. The latter appeared in a total of 9,739 protein-coding genes. The median number of non-synonymous substitutions per sample was 18 (MAD = 19).

Editing ratios had a long-tailed skewed distribution (median = 0.14), suggesting that these variants are present in a small number of transcripts. Most of the non-synonymous sites (90.8%, 46,245/50,883) were found in single samples. Thus, although variants are pervasive on a global scale, many reflect background noise.

To focus on likely functional sites, we looked for conserved non-synonymous substitutions, that were present in >20 tissue samples and had a median editing ratio >0.1 in any tissue. Using these criteria, we found 65 conserved editing sites in 44 genes (Table S4). Six sites were predicted to be deleterious according to PROVEAN (*Choi et al., 2012*) and SIFT (*Ng & Henikoff, 2001*) (chr4:57,110,146, chr6:44,152,612, chr9:33,271,197, chr16:5,044,722, chr16:57,683,958 and chr17:1,534,142). As expected, conserved editing sites had higher editing ratios (median$_{conserved}$ = 0.30, median$_{notconserved}$ = 0.05, $P < 2.2e-16$, Mann–Whitney U test). Moreover, 80% (52/65) of these sites had been reported in REDIportal (*Picardi et al., 2017*), and 35% (23/65) had been reported in the literature (Table S4). The classical example of recoding in *GRIA2* (glutamate ionotropic receptor AMPA type subunit 2) was found at the known site (chr4:157,336,723; p.Q607R) (*Wright & Vissel, 2012*) in aortic wall and internal mammary artery. The median editing ratio in each tissue was 1, indicating that all transcripts carried the modified base.

Eleven recoding sites were novel (Table 2). Seven had low editing ratios (<0.5; median = 0.24), which in part may explain why they were not detected previously. This finding highlights the power of large RNA-seq datasets such as STARNET.

Table 2 Novel A-to-G(I) mRNA-recoding sites.

Site[a]	Gene	Amino acid change	Tissue(s)[b]	Median editing ratio[c]	N[d]
4:57,110,146	IGFBP7	E69G	AOR, MAM	0.11	197
7:38,262,191	TARP	N58S	BLO	0.40	37
8:144,247,668	MROH1	S1037G, S1028G	LIV, VAF	0.47	94
9:33,271,197	CHMP5	K121E	BLO, SUF	0.07	853
10:45,789,442	FAM21C	K1158R, K1124R, K1199R	BLO, LIV, SUF, SKM, VAF	0.50	216
12:57,625,434	SLC26A10	R500G	MAM	0.29	74
16:30,188,879	CORO1A	E434G	VAF	0.17	22
17:1,534,142	PITPNA	D242G	LIV	0.08	55
19:7,520,407	ZNF358	S389G	MAM	0.09	95
19:54,221,229	LILRB3	Q270R	FOC, MAC	0.21	251
19:54,221,256	LILRB3	E261G	FOC, MAC	0.50	223

Notes.

[a]Genomic coordinate of the site (GRCh38). Given as chromosome:position.

[b]Tissue abbreviations: aortic wall (AOR), internal mammary artery (MAM), whole blood (BLO), liver (LIV), visceral fat (VAF), subcutaneous fat (SUF), skeletal muscle (SKM), foam cells (FOC), and macrophages (MAC).

[c]Across all samples.

[d]Number of samples in which editing was found. This number may include samples whose editing signal was lower than required to meet the criteria for detection.

Two recoding sites (one previously reported *Zhang et al., 2014a* and one novel) resided in *IGFBP7* (insulin-like growth factor-binding protein 7). The first site, chr4:57,110,062 changes a codon encoding lysine to arginine (amino acid 97); it was found specific to the internal mammary artery and had been reported in brain (*Zhang et al., 2014a*). The second, chr4:57,110,146, changes a codon encoding glutamic acid to glycine (amino acid 69); it was found in both atherosclerotic aortic wall and the internal mammary artery. Both recoding sites were in the IGF-binding domain of IGFBP7, possibly highlighting a novel regulatory mechanism.

Two recoding sites (chr2:219,483,602 in *SPEG* and chr3:179,375,226 in *MFN1*) were previously reported only in mouse (*Danecek et al., 2012*), suggesting these mRNA-recoding sites are of ancient origin since they were present in the common ancestor of extant human and mouse lineages (diverged ∼65 to 110 million years ago (*Emes et al., 2003*)).

Next, we investigated A-to-I recoding events resulting in the loss of stop codons (the opposite, the gaining of stop codons, is not biologically possible by A-to-I editing). Using the same criteria described above, we found one recoding site, at chr10:101,017,585 in *PDZD7* (PDZ domain containing 7). It was specific to visceral abdominal fat and modifies the stop codon UAG to UGG (encodes tryptophan), extending PDZD7 by 18 amino acids (WSQIPGLKLSTRLSLPKC). This recoding site was recently reported in human brain tissue (*Hwang et al., 2016*). To our knowledge this is the first and perhaps only recoding site in humans that causes protein extension. We checked if the change was associated with measured traits (e.g., LDL-C) and we did not find any association.

Table 3 A-to-G(I) RNA editing sites associated with clinical parameters.

Site[a]	Tissue[b]	Trait	P-value	FDR[c]	Gene	Region
1:204,556,653	LIV	HDL	2.7e−06	0.04	MDM4	3′ UTR
2:37,100,523	BLO	LDL	5.1e−06	0.03	EIF2AK2	3′ UTR
6:149,822,432	SUF	LDL	7.0e−06	0.01	LRP11	Intron
10:50,804,961	LIV	LDL/HDL	5.7e−07	0.01	A1CF	3′ UTR
12:7,097,577	LIV	LDL/HDL	5.9e−07	0.01	C1RL	Intron
12:68,843,961	LIV	18:2, linoleic acid	8.2e−08	0.003	MDM2	3′ UTR
19:4,526,500	LIV	HDL	3.1e−07	0.005	PLIN5	Intron
19:6,683,589	LIV	HDL	2.7e−06	0.04	C3	Intron, RE[d]
19:11,449,682	LIV	VLDL	1.4e−06	0.02	PRKCSH	Intron

Notes.
[a] Genomic coordinate of the site (GRCh38). Given as chromosome:position.
[b] Tissue abbreviations: aortic wall (AOR), internal mammary artery (MAM), whole blood (BLO), liver (LIV), visceral fat (VAF), subcutaneous fat (SUF), skeletal muscle (SKM), foam cells (FOC), and macrophages (MAC).
[c] False discovery rate (permutation P-value). Based on 10,000 permutations.
[d] Regulatory element (RE).

pri-mir-27b and mir-605 editing in aortic wall and internal mammary artery are associated with increased plasma levels of high-density lipoprotein

In preparing the sequencing libraries, we used the Ribo-Zero protocol to eliminate ribosomal RNA, which enabled us to investigate small non-coding RNAs in 996 RNA-seq samples from atherosclerotic aortic wall and internal mammary artery. We identified A-to-G(I) editing sites in 102 unique microRNAs (miRNAs; annotations span the entire predicted stem-loop region) and 73 small nucleolar RNAs (snoRNAs). Stringent editing criteria (recurrence in > 50 samples) identified 18 editing sites in 10 miRNAs (mir-24-2, mir-27a, mir-27b, mir-605, mir-641, mir-1254, mir-1285-1, mir-1304, mir-3126, and mir-3138; Table S5) and 2 snoRNAs (U88 and SNORD17). RNA editing sites in mir-24-2, mir-605, mir-3126, and U88 had not previously been reported in REDIportal (*Picardi et al., 2017*) or DARNED (*Kiran & Baranov, 2010*). Six editing sites occurred in mature miRNAs: three in mir-605-3p and one each in mir-3138, mir-605-5p, and mir-24-2-3p. Two editing sites were associated with higher HDL cholesterol levels: chr9:95,085,457, which was located outside the mature miRNA (pri-mir-27b; $P = 2.2e−05$; FDR = 8.5%) and chr10:51,299,636 (mir-605-3p; $P = 1.6e−06$; FDR = 0.65%).

RNA editing in 3′ UTRs of several genes is associated with plasma lipid levels

Next, we looked for associations between 18,458 RNA editing sites and 208 clinical parameters (*Franzén et al., 2016*). Thirty-nine associations were found significant (FDR < 5%; 10,000 permutations) and involved 25 genes in five tissues (Table S6). 52% (13/25) of these were located in 3′ UTRs, implying they could affect transcript regulation. Of the remaining sites, eight were in introns, two were in known regulatory elements, one was exonic, and three were intergenic. A subset of all significant associations is listed in Table 3.

Although none of the 25 genes had been directly associated with coronary artery disease, 22 of 25 sites were associated with plasma lipoprotein levels. Two of the affected genes may
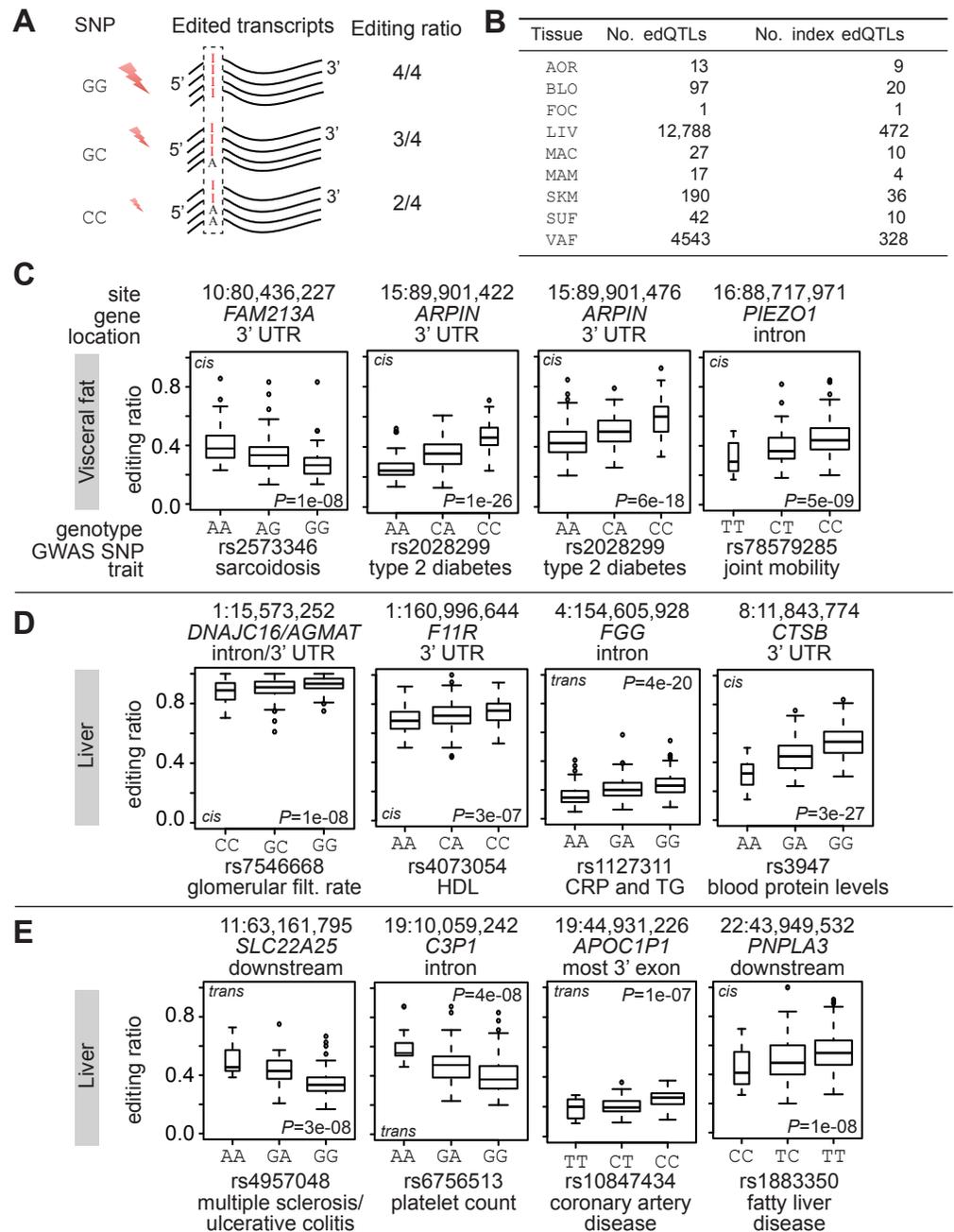
Figure 3 **RNA editing quantitative trait loci.** (A) Schematic illustration of an RNA editing QTL (edQTL). The genomic marker (e.g., a SNP) is associated with RNA editing levels at a specific site in a transcript. (B) Summary of number of identified (index) edQTLs per tissue. (C–E) Box plots of edQTLs where regulatory SNPs coincide with lead SNPs from GWAS. *P*-values refer to non-rank-normal transformed data. The genomic coordinate of the RNA editing site and the overlapping gene are shown on top of each plot. The GWAS SNP is shown at the bottom together with its reported trait. The width of the box is proportional to the number of observations it contains. Rings are outlier samples. The molecular distance is indicated on the top of each subplot: (continued on next page...)

Figure 3 (…continued)
(*cis*) intra-chromosome association; and (*trans*) inter-chromosome association. GWAS *P*-values: rs2573346 (sarcoidosis), $P = 9e-16$; rs2028299 (type 2 diabetes), $P = 1e-11$; rs78579285 (joint mobility), $P = 6e-12$; rs7546668 (glomerular filtration rate), $P = 1e-09$; rs4073054 (HDL), $P = 5e-11$; rs1127311 (C-reactive protein or triglycerides), $P = 6e-09$; rs3947 (blood protein levels), $P = 2e-27$; rs4957048 (multiple sclerosis/ulcerative colitis), $P = 1e-09$; rs6756513 (platelet count), $P = 7e-10$; rs10847434 (coronary artery disease), $P = 6e-15$; and rs1883350 (fatty liver disease), $P = 4e-10$.

have roles in lipid metabolism: *LRP11* (low density lipoprotein receptor-related protein 11) and *PLIN5* (lipid storage droplet protein 5). In subcutaneous fat, the *LRP11* editing site (chr6:149,822,432), located in the most 3′ intron of the longer isoform, was associated with plasma low density lipoprotein (LDL) levels. RNA editing of two genes known for their roles in innate immune response—*C3* (complement component 3) and *C1RL* (complement C1r subcomponent like)—occurred in liver and were both associated with plasma HDL levels.

QTL mapping of RNA editing identifies novel candidate disease genes

To uncover possible genetic regulation of RNA editing (*Kurmangaliyev, Ali & Nuzhdin, 2015*), we looked for RNA editing quantitative trait loci (edQTLs; Fig. 3A). Using data from the 1000 Genome Project (*Durbin et al., 2010*), imputation resulted in 6,246,842 common variants with minor allele frequency >5%. Next, we identified 17,718 edQTLs at FDR <20% (i.e., associations between allele dosages and RNA editing ratios; Fig. 3B) by performing ~2.6 billion hypothesis tests (comprising 11,280 unique editing sites). We examined whether identified edQTLs were also expression QTLs (eQTLs). 19.5% (3,465/17,718) of RNA editing sites in edQTLs were eQTLs with respect to the overlapping gene. Across all tissues, we found 890 index edQTLs—the most significant association for each RNA editing site (Dataset S2). A total of 77% (690/890) of the index edQTLs were *trans* (the regulatory SNP and the editing site are located on different chromosomes).

Most of the editing sites with edQTLs were enriched in *Alu* elements (87%, 783/890); the *Alu*-subtype distribution was the same as shown in Fig. 1F. A total of 95% (853/890) of the edQTLs have previously reported editing sites (*Kiran & Baranov, 2010*; *Picardi et al., 2017*); 47% (425/890) were in 3′ UTRs, 25% (230/890) were in introns, and 12% (113/890) were in intergenic regions. RNA editing sites of edQTLs overlapped 417 genes (Dataset S2).

To assess a possible mechanistic role of edSNPs in disease, we intersected edSNPs with lead SNPs from GWAS. 90 edSNPs coincided with GWAS SNPs, encompassing 59 traits. LD-pruning collapsed the 90 SNPs to 30 non-linked SNPs. rs2573346 is associated with sarcoidosis (*Hofmann et al., 2008*) and had an edQTL in the 3′ UTR of *FAM213A* in visceral fat (Fig. 3C). rs2028299 is associated with type 2 diabetes (*Kooner et al., 2011*) and had two edQTLs in visceral fat (both located in the 3′ UTR of *ARPIN*). rs2028299 is associated with expression of *ARPIN* in fat tissue ($P < 1e-10$) (*Kooner et al., 2011*). A negative correlation was found between *ARPIN* expression in visceral fat and plasma levels of VLDL ($r = -0.2$; $P < 1e-06$). rs78579285 has been associated with joint mobility (Beighton score) (*Pickrell et al., 2016*) and we detected an edQTL in visceral fat within the intron of *PIEZO1* (encodes a mechanosensitive ion-channel component). rs7546668 is associated with kidney function

(glomerular filtration rate) (*Gorski et al., 2017*) and it formed an edQTL with an editing site detected in liver (Fig. 3D). The editing site overlapped *DNAJC16* (encodes DnaJ Heat Shock Protein Family (Hsp40) Member C16) and *AGMAT* (encodes Agmatinase). rs4073054 is associated with HDL (*Chasman et al., 2009*) and had an edQTL with an RNA editing site in the 3′ UTR of *F11R* in liver. Expression of *F11R* was associated with plasma levels of VLDL ($r = -0.2$, $P < 1e-08$) and HDL ($r = -0.2$, $P < 1e-07$). rs1127311 is a pleiotropic locus and it is associated with C-reactive protein and triglyceride levels (*Ligthart, Vaez & Hsu, 2016*). Interestingly, rs1127311 is situated in the 3′ UTR of *ADAR*, and it forms an edQTL with an editing site in the second to the last intron of *FGG*. rs3947, situated in the 3′ UTR of the protease-encoding gene *CTSB*, is associated with blood protein levels (*Suhre et al., 2017*). rs3947 forms an edQTL with an editing site located in the 3′ UTR of *CTSB*. rs3947 is in strong LD with rs2740594 (R-sq = 0.86), which is associated with Parkinson's disease (*Chang et al., 2017*). rs4957048 is associated with multiple sclerosis/ulcerative colitis (*De Lange et al., 2017*), and it was associated with an editing site located downstream of *SLC22A25* (Fig. 3E). rs6756513 is associated with platelet count (*Astle et al., 2016*) and it had an edQTL in an intron of *C3P1*. rs10847434 is associated with coronary artery disease (*Lee et al., 2013*) and had an edQTL with an editing site in the most 3′ exon of *APOC1P1* in liver. *APOC1P1* encodes a long non-coding RNA without any known function or role. However, its locus (*Jeemon et al., 2011*) has been extensively linked to coronary artery disease (neighboring genes are *APOE*, *APOC1*, and *TOMM40*). In liver, there was a negative correlation between expression of *APOC1P1* and *APOE* ($r = -0.3$; $P = 7e-13$) as well as *APOC1P1* and *APOC1* ($r = -0.22$; $P = 1e-07$). The negative correlation may indicate a regulatory function of *APOC1P1*. rs1883350 is associated with fatty liver disease (*Feitosa et al., 2013*) and we identified one liver edQTL slightly downstream of *PNPLA3* (Patatin-like phospholipase domain-containing protein 3). *PNPLA3* is associated with fatty liver disease (*Speliotes et al., 2010*) and encodes a protein known to be involved in liver fat storage. Finally, rs4739066 is associated with myocardial infarction (*Wakil et al., 2016*) and we found this SNP to have two edQTLs with editing sites in the 3′ UTR of the α-tocopherol transfer protein gene (*TTPA*). The editing sites are located 64 bp apart and are intriguing since they have opposite effect sizes. *TTPA* is associated with the severity of atherosclerotic lesions in the proximal aorta (*Terasawa et al., 2000*).

DISCUSSION

We analyzed RNA editing by revisiting RNA-seq data from the STARNET study (*Franzén et al., 2016*). The study aimed to characterize the RNA editing landscape of human tissues and link RNA editing events with phenotypic traits. We analyzed several thousand RNA-seq samples using a novel detection pipeline for RNA editing. Our analysis covered 855 subjects, seven human tissues and two cell types. We identified ~1.6 million RNA editing events, representing a strong A-to-I editing signal. The identified events constituted novel editing sites and previously reported sites (*Picardi et al., 2017*). Many novel events originated from intergenic regions and non-coding RNAs, indicating that many transcripts of the human genome are still uncharacterized. The A-to-I editing levels ranged from almost undetectable

to 100%, suggesting that unidentified factors regulate the strength of editing at individual sites. As expected, the second type of canonical editing, C-to-U, was much less common (~3.5% of identified events). Similar to previous studies (*Park et al., 2012*; *Ramaswami et al., 2012*; *Ramaswami et al., 2013*) we identified a large number of noncanonical edits, which may be true editing events, ultra-rare genetic variants or artifacts from sequencing and data processing. Noncanonical U-to-C was the second most common event (Fig. 1B) as shown by analysis of strand-specific data, confirming a previous report (*Bahn et al., 2012*). This finding may point to a novel editing mechanism in humans. The existence of putative U-to-C editing remains uncertain, and informatics alone is unlikely to settle this question.

We also found RNA editing to be strongly enriched in *Alu* elements, as previously reported (*Athanasiadis, Rich & Maas, 2004*; *Kim et al., 2004*; *Ramaswami et al., 2012*). Within and outside of *Alu* elements the pattern of editing was stochastic, reflecting the lack of specificity of the RNA editing machinery. The relationship between the number of events and sequencing depth also suggested that editing is unspecific, since there was no tendency to saturation. Most of the events were tissue-specific, likely reflecting differences in gene expression between tissues. Functional RNA editing sites have been reported in several studies, e.g., (*Sommer et al., 1991*). We took advantage of our multi-sample cohort to identify functional sites. Remarkably, across 1.6 million RNA editing sites, we discovered only 11 new sites that alter the amino acid sequence. Such recoding sites were inferred from their occurrence in multiple subjects. Sequence analysis revealed that most of them were neutral, indicating the role of purifying selection in removing harmful sites. Notably, *GRIA2* mRNAs were edited in aortic wall and internal mammary artery. *GRIA2* editing converts a codon for glutamine to arginine, which renders ion-channels impermeable to calcium ions. *GRIA2* is ubiquitously expressed in brain tissues as well as in the aortic wall, fallopian tubes, pituitary gland, and uterus (*Melé et al., 2015*). The role of these receptors outside the central nervous system is intriguing yet unknown.

We found that 39 RNA editing sites were associated with traits, raising the possibility of phenotypic consequences. In most cases editing was found in the 3′ UTR or an intron. The former is a target for transcript regulation, and may suggest disruption of binding sites of miRNAs or RNA-binding proteins. Several sites were found in introns, without obvious evidence for perturbed regulatory elements. While we are not able to rule out that some of these associations may have arisen due to unmeasured confounding factors, we believe further investigation is warranted.

Several A-to-I editing events were in important miRNA families, possibly as a mechanism to regulate miRNA function through modification of binding specificity. We found an association between plasma HDL levels and editing of pri-mir-27b/mir-605. mir-27b has been linked to progression of atherosclerosis (*Li et al., 2011a*; *Zhang et al., 2014b*), and mir-605 is implicated in stroke (*Yuan et al., 2016*) and hypertension (*Li et al., 2011b*). These data highlight the need to further integrate A-to-I editing data with respect to the short non-coding transcriptome. More specialized sequencing assays will be needed to capture the full extent of A-to-I editing in miRNAs and other small non-coding RNAs.

Finally, to examine the genetic regulation of RNA editing, we generated and examined edQTLs. edQTLs link RNA editing to genetic loci and have been described in *Drosophila melanogaster* (*Kurmangaliyev, Ali & Nuzhdin, 2015*; *Ramaswami et al., 2015*) and humans (*Park et al., 2017*). Intersecting regulatory SNPs with known disease-associated variants revealed putative causal links for several GWAS SNPs. We found 30 disease-associated regulatory SNPs, several of which have been linked to cardiometabolic traits. For example, rs2028299, previously associated with type 2 diabetes, was an edQTL in the transcript of *ARPIN*, suggesting dysregulation of RNA editing as the cause of increased disease risk.

CONCLUSION

Our data indicate that RNA editing events are extremely common, but rarely of biological significance. Nevertheless, in certain instances such as among single nucleotide polymorphisms from genome-wide association studies, and for specific genes in disease-relevant tissues, RNA editing appears likely to play an important and potentially causal role in disease pathobiology. These findings highlight a novel mechanism of genetic contribution to cardiometabolic disease, and suggest that further investigation of this process is warranted.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Oscar Franzén conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Raili Ermel, Katyayani Sukhavasi, Rajeev Jain and Anamika Jain performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Christer Betsholtz, Chiara Giannarelli, Jason C. Kovacic, Arno Ruusalepp, Josefin Skogsberg, Ke Hao, Eric E. Schadt and Johan L.M. Björkegren conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

Ethical approvals and informed consents were obtained as previously described (Franzén et al. Science 2016; Karolinska Institutet Dnr 154/7 and 188/M-12).

Data Availability

The following information was supplied regarding data availability:

Code is available in GitHub: https://github.com/oscar-franzen/rnaed.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.4466#supplemental-information.

REFERENCES

Anders S, Pyl PT, Huber W. 2014. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166–169 DOI 10.1093/bioinformatics/btu638.

Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, Lambourne JJ, Sivapalaratnam S, Downes K, Kundu K, Bomba L, Berentsen K, Bradley JR, Daugherty LC, Delaneau O, Freson K, Garner SF, Grassi L, Guerrero J, Haimel M, Janssen-Megens EM, Kaan A, Kamat M, Kim B, Mandoli A, Marchini J, Martens JHA, Meacham S, Megy K, O'Connell J, Petersen R, Sharifi N, Sheard SM, Staley JR, Tuna S, Van der Ent M, Walter K, Wang S-Y, Wheeler E, Wilder SP, Iotchkova V, Moore C, Sambrook J, Stunnenberg HG, Di Angelantonio E, Kaptoge S, Kuijpers TW, Carrillo-de Santa-Pau E, Juan D, Rico D, Valencia A, Chen L, Ge B, Vasquez L, Kwan T, Garrido-Martín D, Watt S, Yang Y, Guigo R, Beck S, Paul DS, Pastinen T, Bujold D, Bourque G, Frontini M, Danesh J, Roberts DJ, Ouwehand WH, Butterworth AS, Soranzo N. 2016. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**:1415–1429 e19 DOI 10.1016/j.cell.2016.10.042.

Athanasiadis A, Rich A, Maas S. 2004. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLOS Biology* **2(12)**:e391 DOI 10.1371/journal.pbio.0020391.

Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**:1294–1296 DOI 10.1093/bioinformatics/btm108.

Bahn JH, Lee J-H, Li G, Greer C, Peng G, Xiao X. 2012. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Research* **22**:142–150 DOI 10.1101/gr.124107.111.

Bass BL. 2002. RNA editing by adenosine deaminases that act on RNA. *Annual Review of Biochemistry* **71**:817–846 DOI 10.1146/annurev.biochem.71.110601.135501.

Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs FJ, Rechavi G, Li JB, Eisenberg E, Levanon EY. 2014. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Research* **24**:365–376 DOI 10.1101/gr.164749.113.

Benne R, Van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**:819–826.

Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE. 2008. Active Alu retrotransposons in the human genome. *Genome Research* **18**:1875–1883 DOI 10.1101/gr.081737.108.

Björkegren JLM, Kovacic JC, Dudley JT, Schadt EE. 2015. Genome-wide significant loci: how important are they?: systems genetics to understand heritability of coronary artery disease and other common complex disorders. *Journal of the American College of Cardiology* **65**:830–845 DOI 10.1016/j.jacc.2014.12.033.

Blanc V, Davidson NO. 2003. C-to-U RNA editing: mechanisms leading to genetic diversity. *The Journal of Biological Chemistry* **278**:1395–1398 DOI 10.1074/jbc.R200024200.

De Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, Jostins L, Rice DL, Gutierrez-Achury J, Ji S-G, Heap G, Nimmo ER, Edwards C, Henderson P, Mowat C, Sanderson J, Satsangi J, Simmons A, Wilson DC, Tremelling M, Hart A, Mathew CG, Newman WG, Parkes M, Lees CW, Uhlig H, Hawkey C, Prescott NJ, Ahmad T, Mansfield JC, Anderson CA, Barrett JC. 2017. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature Genetics* **49**:256–261 DOI 10.1038/ng.3760.

Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, Van der Brug M, Cai F, International Parkinson's Disease Genomics Consortium, 23andMe Research Team G, Kerchner GA, Ayalon G, Bingol B, Sheng M, Hinds D, Behrens TW, Singleton AB, Bhangale TR, Graham RR. 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nature Genetics* **49**:1511–1516 DOI 10.1038/ng.3955.

Chasman DI, Paré G, Mora S, Hopewell JC, Peloso G, Clarke R, Cupples LA, Hamsten A, Kathiresan S, Mälarstig A, Ordovas JM, Ripatti S, Parker AN, Miletich JP, Ridker PM. 2009. Forty-three loci associated with plasma lipoprotein size, concentration, and cholesterol content in genome-wide analysis. *PLOS Genetics* **5**:e1000730 DOI 10.1371/journal.pgen.1000730.

Chen CX, Cho DS, Wang Q, Lai F, Carter KC, Nishikura K. 2000. A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* **6**:755–767.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. *PLOS ONE* **7**:e46688 DOI 10.1371/journal.pone.0046688.

Danecek P, Nellåker C, McIntyre RE, Buendia-Buendia JE, Bumpstead S, Ponting CP, Flint J, Durbin R, Keane TM, Adams DJ. 2012. High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biology* **13**:26 DOI 10.1186/gb-2012-13-4-r26.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21 DOI 10.1093/bioinformatics/bts635.

Durbin RM, Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Collins FS, De La Vega FM, Donnelly P, Egholm M, Flicek P, Gabriel SB, Gibbs RA, Knoppers BM, Lander ES, Lehrach H, Mardis ER, McVean GA, Nickerson DA, Peltonen L, Schafer AJ, Sherry ST, Wang J, Wilson RK, Gibbs RA, Deiros D, Metzker M, Muzny D, Reid J, Wheeler D, Wang J, Li J, Jian M, Li G, Li R, Liang H, Tian G, Wang B, Wang J, Wang W, Yang H, Zhang X, Zheng H, Lander ES, Altshuler DL, Ambrogio L, Bloom T, Cibulskis K, Fennell TJ, Gabriel SB, Jaffe DB, Shefler E, Sougnez CL, Bentley DR, Gormley N, Humphray S, Kingsbury Z, Koko-Gonzales P, Stone J, McKernan KJ, Costa GL, Ichikawa JK, Lee CC, Sudbrak

R, Lehrach H, Borodina TA, Dahl A, Davydov AN, Marquardt P, Mertes F, Nietfeld W, Rosenstiel P, Schreiber S, Soldatov AV, Timmermann B, Tolzmann M, Egholm M, Affourtit J, Ashworth D, Attiya S, Bachorski M, Buglione E, Burke A, Caprio A, Celone C, Clark S, Conners D, Desany B, Gu L, Guccione L, Kao K, Kebbel A, Knowlton J, Labrecque M, McDade L, Mealmaker C, Minderman M, Nawrocki A, Niazi F, Pareja K, Ramenani R, Riches D, Song W, Turcotte C, Wang S, Mardis ER, Wilson RK, Dooling D, Fulton L, Fulton R, Weinstock G, Durbin RM, Burton J, Carter DM, Churcher C, Coffey A, Cox A, Palotie A, Quail M, Skelly T, Stalker J, Swerdlow HP, Turner D, De Witte A, Giles S, Gibbs RA, Wheeler D, Bainbridge M, Challis D, Sabo A, Yu F, Yu J, Wang J, Fang X, Guo X, Li R, Li Y, Luo R, Tai S, Wu H, Zheng H, Zheng X, Zhou Y, Li G, Wang J, Yang H, Marth GT, Garrison EP, Huang W, Indap A, Kural D, Lee W-P, Fung Leong W, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu J, Lee C, Mills RE, Shi X, Daly MJ, DePristo MA, Altshuler DL, Ball AD, Banks E, Bloom T, Browning BL, Cibulskis K, Fennell TJ, Garimella KV, Grossman SR, Handsaker RE, Hanna M, Hartl C, Jaffe DB, Kernytsky AM, Korn JM, Li H, Maguire JR, McCarroll SA, McKenna A, Nemesh JC, Philippakis AA, Poplin RE, Price A, Rivas MA, Sabeti PC, Schaffner SF, Shefler E, Shlyakhter IA, Cooper DN, Ball E V, Mort M, Phillips AD, Stenson PD, Sebat J, Makarov V, Ye K, et al. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**:1061–1073 DOI 10.1038/nature09534.

Emes RD, Goodstadt L, Winter EE, Ponting CP. 2003. Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Human Molecular Genetics* **12**:701–709.

Erikson GA, Bodian DL, Rueda M, Molparia B, Scott ER, Scott-Van Zeeland AA, Topol SE, Wineinger NE, Niederhuber JE, Topol EJ, Torkamani A. 2016. Whole-genome sequencing of a healthy aging cohort. *Cell* **165**:1002–1011 DOI 10.1016/j.cell.2016.03.022.

Farajollahi S, Maas S. 2010. Molecular diversity through RNA editing: a balancing act. *Trends in Genetics* **26**:221–230 DOI 10.1016/j.tig.2010.02.001.

Feitosa MF, Wojczynski MK, North KE, Zhang Q, Province MA, Carr JJ, Borecki IB. 2013. The ERLIN1-CHUK-CWF19L1 gene cluster influences liver fat deposition and hepatic inflammation in the NHLBI Family Heart Study. *Atherosclerosis* **228**:175–180 DOI 10.1016/j.atherosclerosis.2013.01.038.

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U, Campbell PJ. 2015. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research* **43**:D805–D811 DOI 10.1093/nar/gku1075.

Franzén O, Ermel R, Cohain A, Akers NK, Di Narzo A, Talukdar HA, Foroughi-Asl H, Giambartolomei C, Fullard JF, Sukhavasi K, Köks S, Gan L-M, Giannarelli C, Kovacic JC, Betsholtz C, Losic B, Michoel T, Hao K, Roussos P, Skogsberg J, Ruusalepp A, Schadt EE, Björkegren JLM. 2016. Cardiometabolic risk loci share

downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**:827–830 DOI 10.1126/science.aad6970.

Galeano F, Tomaselli S, Locatelli F, Gallo A. 2012. A-to-I RNA editing: the "ADAR" side of human cancer. *Seminars in Cell & Developmental Biology* **23**:244–250 DOI 10.1016/j.semcdb.2011.09.003.

Gorski M, Van der Most PJ, Teumer A, Chu AY, Li M, Mijatovic V, Nolte IM, Cocca M, Taliun D, Gomez F, Li Y, Tayo B, Tin A, Feitosa MF, Aspelund T, Attia J, Biffar R, Bochud M, Boerwinkle E, Borecki I, Bottinger EP, Chen M-H, Chouraki V, Ciullo M, Coresh J, Cornelis MC, Curhan GC, d'Adamo AP, Dehghan A, Dengler L, Ding J, Eiriksdottir G, Endlich K, Enroth S, Esko T, Franco OH, Gasparini P, Gieger C, Girotto G, Gottesman O, Gudnason V, Gyllensten U, Hancock SJ, Harris TB, Helmer C, Höllerer S, Hofer E, Hofman A, Holliday EG, Homuth G, Hu FB, Huth C, Hutri-Kähönen N, Hwang S-J, Imboden M, Johansson Å, Kähönen M, König W, Kramer H, Krämer BK, Kumar A, Kutalik Z, Lambert J-C, Launer LJ, Lehtimäki T, De Borst M, Navis G, Swertz M, Liu Y, Lohman K, Loos RJF, Lu Y, Lyytikäinen L-P, McEvoy MA, Meisinger C, Meitinger T, Metspalu A, Metzger M, Mihailov E, Mitchell P, Nauck M, Oldehinkel AJ, Olden M, Penninx BWJH, Pistis G, Pramstaller PP, Probst-Hensch N, Raitakari OT, Rettig R, Ridker PM, Rivadeneira F, Robino A, Rosas SE, Ruderfer D, Ruggiero D, Saba Y, Sala C, Schmidt H, Schmidt R, Scott RJ, Sedaghat S, Smith AV, Sorice R, Stengel B, Stracke S, Strauch K, Toniolo D, Uitterlinden AG, Ulivi S, Viikari JS, Völker U, Vollenweider P, Völzke H, Vuckovic D, Waldenberger M, Jin Wang J, Yang Q, Chasman DI, Tromp G, Snieder H, Heid IM, Fox CS, Köttgen A, Pattaro C, Böger CA, Fuchsberger C. 2017. 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific Reports* **7**:45040 DOI 10.1038/srep45040.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, Tress M, Rodriguez JM, Ezkurdia I, Van Baren J, Brent M, Haussler D, Kellis M, Valencia A, Reymond A, Gerstein M, Guigó R, Hubbard TJ. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Research* **22**:1760–1774 DOI 10.1101/gr.135350.111.

Hideyama T, Yamashita T, Suzuki T, Tsuji S, Higuchi M, Seeburg PH, Takahashi R, Misawa H, Kwak S. 2010. Induced loss of ADAR2 engenders slow death of motor neurons from Q/R Site-Unedited GluR2. *Journal of Neuroscience* **30**:11917–11925 DOI 10.1523/JNEUROSCI.2021-10.2010.

Hofmann S, Franke A, Fischer A, Jacobs G, Nothnagel M, Gaede KI, Schürmann M, Müller-Quernheim J, Krawczak M, Rosenstiel P, Schreiber S. 2008. Genome-wide association study identifies ANXA11 as a new susceptibility locus for sarcoidosis. *Nature Genetics* **40**:1103–1106 DOI 10.1038/ng.198.

Hood JL, Emeson RB. 2011. Editing of neurotransmitter receptor and ion channel rnas in the nervous system. *Current Topics in Microbiology and Immunology* **353**:61–90 DOI 10.1007/82_2011_157.

Hwang T, Park C-K, Leung AKL, Gao Y, Hyde TM, Kleinman JE, Rajpurohit A, Tao R, Shin JH, Weinberger DR. 2016. Dynamic regulation of RNA editing in human brain development and disease. *Nature Neuroscience* **19**:1093–1099 DOI 10.1038/nn.4337.

Jeemon P, Pettigrew K, Sainsbury C, Prabhakaran D, Padmanabhan S. 2011. Implications of discoveries from genome-wide association studies in current cardiovascular practice. *World Journal of Cardiologyy* **3**:230–247 DOI 10.4330/wjc.v3.i7.230.

Kim M, Hur B, Kim S. 2016. RDDpred: a condition-specific RNA-editing prediction model from RNA-seq data. *BMC Genomics* **17**:5 DOI 10.1186/s12864-015-2301-y.

Kim DDY, Kim TTY, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A. 2004. Widespread RNA editing of embedded Alu elements in the human transcriptome. *Genome Research* **14**:1719–1725 DOI 10.1101/gr.2855504.

Kiran A, Baranov PV. 2010. DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics* **26**:1772–1776 DOI 10.1093/bioinformatics/btq285.

Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, Been LF, Chia K-S, Dimas AS, Hassanali N, Jafar T, Jowett JBM, Li X, Radha V, Rees SD, Takeuchi F, Young R, Aung T, Basit A, Chidambaram M, Das D, Grundberg E, Hedman Å K, Hydrie ZI, Islam M, Khor C-C, Kowlessur S, Kristensen MM, Liju S, Lim W-Y, Matthews DR, Liu J, Morris AP, Nica AC, Pinidiyapathirage JM, Prokopenko I, Rasheed A, Samuel M, Shah N, Shera AS, Small KS, Suo C, Wickremasinghe AR, Wong TY, Yang M, Zhang F, Abecasis GR, Barnett AH, Caulfield M, Deloukas P, Frayling TM, Froguel P, Kato N, Katulanda P, Kelly MA, Liang J, Mohan V, Sanghera DK, Scott J, Seielstad M, Zimmet PZ, Elliott P, Teo YY, McCarthy MI, Danesh J, Tai ES, Chambers JC, Tai ES, Chambers JC. 2011. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nature Genetics* **43**:984–989 DOI 10.1038/ng.921.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42**:D68–73 DOI 10.1093/nar/gkt1181.

Kurmangaliyev YZ, Ali S, Nuzhdin SV. 2015. Genetic determinants of RNA editing levels of ADAR Targets in *Drosophila melanogaster*. *G3* **6**:391–396 DOI 10.1534/g3.115.024471.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R,

Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng J-F, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, De la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen H-C, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921 DOI 10.1038/35057062.

Lee J-Y, Lee B-S, Shin D-J, Woo Park K, Shin Y-A, Joong Kim K, Heo L, Young Lee J, Kyoung Kim Y, Jin Kim Y, Bum Hong C, Lee S-H, Yoon D, Jung Ku H, Oh I-Y, Kim B-J, Lee J, Park S-J, Kim J, Kawk H-K, Lee J-E, Park H-K, Lee J-E, Nam H-Y, Park H-Y, Shin C, Yokota M, Asano H, Nakatochi M, Matsubara T, Kitajima H, Yamamoto K, Kim H-L, Han B-G, Cho M-C, Jang Y, Kim H-S, Euy Park J, Lee J-Y. 2013. A genome-wide association study of a coronary artery disease risk variant. *Journal of Human Genetics* **58**:120–126 DOI 10.1038/jhg.2012.124.

Li T, Cao H, Zhuang J, Wan J, Guan M, Yu B, Li X, Zhang W. 2011a. Identification of miR-130a, miR-27b and miR-210 as serum biomarkers for atherosclerosis obliterans. *Clinica Chimica Acta; International Journal of Clinical Chemistry* **412**:66–70 DOI 10.1016/j.cca.2010.09.029.

Li MJ, Liu Z, Wang P, Wong MP, Nelson MR, Kocher JPA, Yeager M, Sham PC, Chanock SJ, Xia Z, Wang J. 2016. GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Research* **44**:D869–D876 DOI 10.1093/nar/gkv1317.

Li S, Zhu J, Zhang W, Chen Y, Zhang K, Popescu LM, Ma X, Bond Lau W, Rong R, Yu X, Wang B, Li Y, Xiao C, Zhang M, Wang S, Yu L, Chen AF, Yang X, Cai J. 2011b. Signature microRNA expression profile of essential hypertension and its novel link to human cytomegalovirus infection. *Circulation* **124**:175–184 DOI 10.1161/CIRCULATIONAHA.110.012237.

Ligthart S, Vaez A, Hsu Y-H, Inflammation Working Group of the CHARGE Consortium, PMI-WG-XCP AG, LifeLines Cohort Study, Stolk R, Uitterlinden AG,

Hofman A, Alizadeh BZ, Franco OH, Dehghan A. 2016. Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. *BMC Genomics* **17**:443 DOI 10.1186/s12864-016-2712-4.

Macbeth MR, Schubert HL, Vandemark AP, Lingam AT, Hill CP, Bass BL. 2005. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309**:1534–1539 DOI 10.1126/science.1113150.

Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segrè AV, Djebali S, Niarchou A, GTEx Consortium, Wright FA, Lappalainen T, Calvo M, Getz G, Dermitzakis ET, Ardlie KG, Guigó R. 2015. Human genomics. The human transcriptome across tissues and individuals. *Science* **348**:660–665 DOI 10.1126/science.aaa0355.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**:621–628 DOI 10.1038/nmeth.1226.

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. *Genome Research* **11**:863–874 DOI 10.1101/gr.176601.

Nishikura K. 2010. Functions and regulation of RNA editing by ADAR deaminases. *Annual Review of Biochemistry* **79**:321–349 DOI 10.1146/annurev-biochem-060208-105251.

Nishikura K. 2016. A-to-I editing of coding and non-coding RNAs by ADARs. *Nature Reviews. Molecular Cell Biology* **17**:83–96 DOI 10.1038/nrm.2015.4.

Park E, Guo J, Shen S, Demirdjian L, Wu YN, Lin L, Xing Y. 2017. Population and allelic variation of A-to-I RNA editing in human transcriptomes. *Genome Biology* **18**:Article 143 DOI 10.1186/s13059-017-1270-7.

Park E, Williams B, Wold BJ, Mortazavi A. 2012. RNA editing in the human ENCODE RNA-seq data. *Genome Research* **22**:1626–1633 DOI 10.1101/gr.134957.111.

Paz-Yaacov N, Bazak L, Buchumenski I, Porath HT, Danan-Gotthold M, Knisbacher BA, Eisenberg E, Levanon EY. 2015. Elevated RNA editing activity is a major contributor to transcriptomic diversity in tumors. *Cell Reports* **13**:267–276 DOI 10.1016/j.celrep.2015.08.080.

Picardi E, D'Erchia AM, Lo Giudice C, Pesole G. 2017. REDIportal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Research* **45**:D750–D757 DOI 10.1093/nar/gkw767.

Picardi E, D'Erchia AM, Montalvo A, Pesole G. 2015. Using REDItools to detect RNA editing events in NGS datasets. *Current Protocols in Bioinformatics* **2015**:12121–121215 DOI 10.1002/0471250953.bi1212s49.

Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. 2016. Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics* **48**:709–717 DOI 10.1038/ng.3570.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome

association and population-based linkage analyses. *American Journal of Human Genetics* **81**:559–575 DOI 10.1086/519795.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842 DOI 10.1093/bioinformatics/btq033.

R Core Team. 2015. R: a language and environment for statistical computing. Version 3.2.2. Vienna: the R Foundation for Statistical Computing. *Available at* https://www. R-project.org/.

Ramaswami G, Deng P, Zhang R, Anna Carbone M, Mackay TFC, Billy Li J. 2015. Genetic mapping uncovers cis-regulatory landscape of RNA editing. *Nature Communications* **6**:Article 8194 DOI 10.1038/ncomms9194.

Ramaswami G, Li JB. 2016. Identification of human RNA editing sites: A historical perspective. *Methods* **107**:42–47 DOI 10.1016/j.ymeth.2016.05.011.

Ramaswami G, Lin W, Piskol R, Tan MH, Davis C, Li JB. 2012. Accurate identification of human Alu and non-Alu RNA editing sites. *Nature Methods* **9**:579–581 DOI 10.1038/nmeth.1982.

Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'Connell MA, Li JB. 2013. Identifying RNA editing sites using RNA sequencing data alone. *Nature Methods* **10**:128–132 DOI 10.1038/nmeth.2330.

Shabalin AA. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**:1353–1358 DOI 10.1093/bioinformatics/bts163.

Slotkin W, Nishikura K. 2013. Adenosine-to-inosine RNA editing and human disease. *Genome Medicine* **5**:Article 105 DOI 10.1186/gm508.

Sommer B, Köhler M, Sprengel R, Seeburg PH. 1991. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**:11–19 DOI 10.1016/0092-8674(91)90568-J.

Speliotes EK, Butler JL, Palmer CD, Voight BF, GIANT Consortium, MIGen Consortium, NASH CRN, Hirschhorn JN. 2010. PNPLA3 variants specifically confer increased risk for histologic nonalcoholic fatty liver disease but not metabolic disease. *Hepatology* **52**:904–912 DOI 10.1002/hep.23768.

Stellos K, Gatsiou A, Stamatelopoulos K, Perisic Matic L, John D, Lunella FF, Jaé N, Rossbach O, Amrhein C, Sigala F, Boon RA, Fürtig B, Manavski Y, You X, Uchida S, Keller T, Boeckel J-N, Franco-Cereceda A, Maegdefessel L, Chen W, Schwalbe H, Bindereif A, Eriksson P, Hedin U, Zeiher AM, Dimmeler S. 2016. Adenosine-to-inosine RNA editing controls cathepsin S expression in atherosclerosis by enabling HuR-mediated post-transcriptional regulation. *Nature Medicine* **22**:1140–1150 DOI 10.1038/nm.4172.

Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, Sarwath H, Thareja G, Wahl A, DeLisle RK, Gold L, Pezer M, Lauc G, El-Din Selim MA, Mook-Kanamori DO, Al-Dous EK, Mohamoud YA, Malek J, Strauch K, Grallert H, Peters A, Kastenmüller G, Gieger C, Graumann J. 2017. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nature Communications* **8**:Article 14357 DOI 10.1038/ncomms14357.

Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, Zhang R, Ramaswami G, Ariyoshi K, Gupte A, Keegan LP, George CX, Ramu A, Huang N, Pollina EA, Leeman DS, Rustighi A, Goh YPS, GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration & Visualization—EBI, Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz, Chawla A, Del Sal G, Peltz G, Brunet A, Conrad DF, Samuel CE, O'Connell MA, Walkley CR, Nishikura K, Li JB. 2017. Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**:249–254 DOI 10.1038/nature24041.

Terasawa Y, Ladha Z, Leonard SW, Morrow JD, Newland D, Sanan D, Packer L, Traber MG, Farese RV. 2000. Increased atherosclerosis in hyperlipidemic mice deficient in alpha -tocopherol transfer protein and vitamin E. *Proceedings of the National Academy of Sciences of the United States of America* **97**:13830–13834 DOI 10.1073/pnas.240462697.

Wakil SM, Ram R, Muiya NP, Mehta M, Andres E, Mazhar N, Baz B, Hagos S, Alshahid M, Meyer BF, Morahan G, Dzimiri N. 2016. A genome-wide association study reveals susceptibility loci for myocardial infarction/coronary artery disease in Saudi Arabs. *Atherosclerosis* **245**:62–70 DOI 10.1016/j.atherosclerosis.2015.11.019.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**:e164 DOI 10.1093/nar/gkq603.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* **42**:D1001–D1006 DOI 10.1093/nar/gkt1229.

Wright A, Vissel B. 2012. The essential role of AMPA receptor GluR2 subunit RNA editing in the normal and diseased brain. *Frontiers in Molecular Neuroscience* **5**:Article 34 DOI 10.3389/fnmol.2012.00034.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**:873–881 DOI 10.1093/bioinformatics/btq057.

Yuan Y, Kang R, Yu Y, Liu J, Zhang Y, Shen C, Wang J, Wu P, Shen C, Wang Z. 2016. Crosstalk between miRNAs and their regulated genes network in stroke. *Scientific Reports* **6**:20429 DOI 10.1038/srep20429.

Zhang R, Li X, Ramaswami G, Smith KS, Turecki G, Montgomery SB, Li JB. 2014a. Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nature Methods* **11**:51–54 DOI 10.1038/nmeth.2736.

Zhang M, Wu J-F, Chen W-J, Tang S-L, Mo Z-C, Tang Y-Y, Li Y, Wang J-L, Liu X-Y, Peng J, Chen K, He P-P, Lv Y-C, Ouyang X-P, Yao F, Tang D-P, Cayabyab FS, Zhang D-W, Zheng X-L, Tian G-P, Tang C-K. 2014b. MicroRNA-27a/b regulates cellular cholesterol efflux, influx and esterification/hydrolysis in THP-1 macrophages. *Atherosclerosis* **234**:54–64 DOI 10.1016/j.atherosclerosis.2014.02.008.

Zhang Q, Xiao X. 2015. Genome sequence-independent identification of RNA editing sites. *Nature Methods* **12**:347–350 DOI 10.1038/nmeth.3314.