Human C-to-U Coding RNA Editing Is Largely Nonadaptive

Zhen Liu1,2 and Jianzhi Zhang*2
1State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
2Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI
*Corresponding author: E-mail: jianzhi@umich.edu.
Associate editor: Meredith Yeager

Abstract
C-to-U RNA editing enzymatically converts the base C to U in RNA molecules and could lead to nonsynonymous changes when occurring in coding regions. Hundreds to thousands of coding sites were recently found to be C-to-U edited or editable in humans, but the biological significance of this phenomenon is elusive. Here, we test the prevailing hypothesis that nonsynonymous editing is beneficial because it provides a means for tissue- or time-specific regulation of protein function that may be hard to accomplish by mutations due to pleiotropy. The adaptive hypothesis predicts that the fraction of sites edited and the median proportion of RNA molecules edited (i.e., editing level) are both higher for nonsynonymous than synonymous editing. However, our empirical observations are opposite to these predictions. Furthermore, the frequency of nonsynonymous editing, relative to that of synonymous editing, declines as genes become functionally more important or evolutionarily more constrained, and the nonsynonymous editing level at a site is negatively correlated with the evolutionary conservation of the site. Together, these findings refute the adaptive hypothesis; they instead indicate that the reported C-to-U coding RNA editing is mostly slightly deleterious or neutral, probably resulting from off-target activities of editing enzymes. Along with similar conclusions on the more prevalent A-to-I editing and m6A modification of coding RNAs, our study suggests that, at least in humans, most events of each type of posttranscriptional coding RNA modification likely manifest cellular errors rather than adaptations, demanding a paradigm shift in the research of posttranscriptional modification.

Key words: APOBEC, deleterious, cellular error, neutral, nonsynonymous, synonymous.

Introduction
RNA editing refers to a heterogeneous group of posttranscriptional modifications of RNAs that include insertion, deletion, and nucleotide conversions but exclude common forms of processing (e.g., splicing, 5′-capping, and 3′-polyadenylation) and conversions to nonstandard nucleotides (e.g., m6A) (Covello and Gray 1993; Nishikura 2006; Farajollahi and Maas 2010). The most prevalent editing of mRNAs transcribed from the animal nuclear genome is A-to-I (inosine) editing (Nishikura 2016). For instance, in humans, millions of sites undergo A-to-I editing, although the vast majority of these sites reside in noncoding regions (Li et al. 2009; Peng et al. 2012; Bazak et al. 2014). The physiological function of A-to-I editing is largely unknown, with evidence for functionality present for only a few incidences (Lomeli et al. 1994; Burns et al. 1997; Seeburg and Hartner 2003; Garrett and Rosenthal 2012; Nishikura 2016). Evolutionary analysis suggests that most observed A-to-I editing events in vertebrate coding regions are neutral or slightly deleterious, likely resulting from promiscuous activities of editing enzymes (Xu and Zhang 2014). However, analyses performed in fruit flies, squids, and octopuses suggest that RNA editing has been subjected to positive Darwinian selection in these lineages, although the selective agents remain unclear (Alon et al. 2015; Duan et al. 2017; Liscovitch-Brauer et al. 2017).

Evolutionary studies of RNA editing of animal nuclear genes have been limited to A-to-I editing, because other types of RNA editing are thought to be rare and hence not amenable to meaningful statistical analysis. A prime example is C-to-U editing, which is predominantly catalyzed by AID (activation-induced cytidine deaminase)/APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide) proteins, encoded by a family of ten genes (APOBEC1, 2, 3, A–D, 3 F–H, and 4) in the human genome (Smith et al. 2012). Although C-to-U is the second most common RNA editing of animal nuclear genes and has been known for over 30 years (Chen et al. 1987; Powell et al. 1987), only a few examples are well documented in humans (Vu and Tsukahara 2017). Several recent transcriptomic studies of human C-to-U editing, however, have started to change this situation. Specifically, Sharma et al. (2015) identified ~100 coding sites that are C-to-U edited in M1 macrophages. Furthermore, hypoxia was found to enhance C-to-U RNA editing in monocytes (Baysal et al. 2013), and >100 edited sites in coding regions were detected from hypoxia-induced monocytes (Sharma et al. 2015). Because the relatively high C-to-U editing in macrophages and monocytes is due to the upregulation of cytidine deaminase APOBEC3A (Sharma et al. 2015), Sharma et al. hypothesized that they could identify many edible sites by overexpressing APOBEC3A in any cell line. Indeed, they reported over 2,000 edited coding sites.
upon overexpression of APOBEC3A in the human embryonic kidney cell line HEK293T (Sharma et al. 2017). Although C-to-U editing in HEK293T is nonnatural because APOBEC3A is not normally expressed in this cell line, at least the identified highly edited sites should represent those that are edited in cell types that normally express APOBEC3A strongly.

Despite the progress in transcriptome-wide identification of C-to-U editing, its function remains unclear. Because C-to-U editing can alter the protein sequence when occurring in coding regions, it provides a means for enhancing proteomic diversity without altering the genome, which could be advantageous when different protein functions are needed in different tissues or developmental times. Furthermore, this means of regulation is thought to be particularly important at evolutionarily conserved sites where mutations altering the protein sequence have obviously been selectively purged perhaps due to pleiotropy (Reenan 2005; Gommans et al. 2009).

Analyzing reported human C-to-U editing sites, we here test whether C-to-U editing is generally adaptive, as commonly thought, or is largely nonadaptive, as in the case of human A-to-I editing.

**Results**

**Nonsynonymous Editing Is Less Frequent than Synonymous Editing**

We respectively compiled human C-to-U coding RNA editing data from macrophages during the M1 activation phase, monocytes under hypoxia, and HEK293T cells under APOBEC3A overexpression (table 1). APOBEC3A is known to preferentially edit C when it is preceded with a U (Farajollahi and Maas 2010; Sharma et al. 2015, 2017; Kouno et al. 2017; Shi et al. 2017). Indeed, the vast majority (89.8–97.8%) of the edited sites in coding regions are located in the DNA motif TC (edited C is underlined) (table 1). For this reason, we focus on editing events in TC motifs in all analyses.

We consider a C-to-U editing in coding regions nonsynonymous if it creates a missense or nonsense change; otherwise it is regarded as synonymous. The adaptive hypothesis of C-to-U editing asserts that nonsynonymous editing is beneficial and hence predicts a higher frequency of nonsynonymous editing than synonymous editing, which is expected to be neutral. Among the 89 TC motifs in which C is edited to U in macrophages, editing is nonsynonymous at \( n = 26 \) sites and synonymous at \( s = 63 \) sites. In all human protein coding regions, potential C-to-U editing is nonsynonymous at \( N = 1,171,927 \) TC motifs and synonymous at \( S = 1,006,853 \) TC motifs. Thus, the frequency of nonsynonymous editing observed is \( f_n = n/N = 2.22 \times 10^{-5} \), whereas the frequency of synonymous editing observed is \( f_s = s/S = 6.26 \times 10^{-5} \). The difference between \( f_n \) and \( f_s \) is statistically significant (\( P = 5.53 \times 10^{-6} \), \( \chi^2 \) test; fig. 1A). Using the same method, we calculated \( f_n \) and \( f_s \) in the monocyte and HEK293T data sets, respectively. Consistent with the finding from macrophages, \( f_n \) is significantly lower than \( f_s \) in both monocytes and HEK293T cells (\( P = 0.042 \) and \( 1.39 \times 10^{-14} \), respectively; fig. 1A). These results suggest that genotypes with nonsynonymous C-to-U editing often have reduced fitness and are purged by natural selection such that the observed \( f_n \) is lower than the neutral expectation. We noticed that the fraction of nonsynonymous editing that has been selectively removed \( (1-f_n/f_s) \) is much greater in macrophages (65%) than in monocytes (32%) and HEK293T (27%). This is probably because the condition (M1 activation phase) under which the macrophages were examined is much more physiological than the conditions (hypoxia and APOBEC3A overexpression) under which the other two cell types were respectively examined. Consequently, editing observed in the former has been subjected to natural selection more frequently than editing observed in the latter.

We also examined the \( f_n/f_s \) ratio of individual genes in the three data sets. Because the number of edited sites per gene is very small, many genes show \( f_n > f_s \) simply by chance. However, we found no gene whose \( f_n \) significantly exceeds \( f_s \) at the nominal \( P \) value of 0.05.

**Nonsynonymous Editing Levels Are Lower than Synonymous Editing Levels**

It is possible that some nonsynonymous editing is deleterious and purged by natural selection but the remaining nonsynonymous editing which is observed is advantageous. To test this hypothesis, we examined the editing level of each edited site, which is the fraction of RNA molecules that are edited at the site. The adaptive hypothesis predicts that editing level is generally higher for observed nonsynonymous editing than synonymous editing. Contrary to the prediction, we found the median nonsynonymous editing level \( (L_n) \) lower than the median synonymous editing level \( (L_s) \) in all three cell types (fig. 1B). This difference is statistically significant in macrophages \( (P = 0.025 \), two-tailed Wilcoxon rank sum test) but not in the other two cell types, which again is likely owing to fewer opportunities for natural selection on the observed editing in the latter than the former. Regardless, our results indicate that even observed nonsynonymous editing is generally deleterious rather than beneficial such that its editing level is lower than the neutral expectation reflected by the level of synonymous editing.

**Editing Frequency Reduces with Gene Importance and Evolutionary Constraint**

The above analyses suggest that nonsynonymous C-to-U editing is generally harmful rather than beneficial. Hence, we predict that, relative to \( f_n/f_s \) should decrease as gene importance increases. To verify this prediction, we obtained a list of 183 human essential genes and 14,388 human nonessential genes (see Materials and Methods) and respectively estimated their \( f_n \) and \( f_s \) values. In part due to the small sample sizes of the macrophage and monocyte data sets, no nonsynonymous editing is identified in essential genes. We thus decided to focus on the HEK293T data set, from which we estimated \( f_n/f_s = 0.32 \) for essential genes (fig. 2A). To test whether this value is significantly lower than the corresponding value of nonessential genes, we randomly picked 183 nonessential genes to estimate its \( f_n/f_s \) and repeated this process 1,000 times. Only 3 of the 1,000 \( f_n/f_s \) values from nonessential genes are equal to or lower than the observed...
Evolutionarily Conserved Sites Tend to Have Lower Editing Levels

As mentioned, the adaptive hypothesis of nonsynonymous RNA editing proposes that nonsynonymous editing should preferentially occur at evolutionarily conserved or mutation-intolerant sites because editing is especially less pleiotropic and hence more beneficial than DNA-level changes at such sites (Reenan 2005; Gommans et al. 2009). This hypothesis predicts that 1) evolutionary conservation is greater for sites undergoing nonsynonymous editing than those undergoing synonymous editing; and 2) editing level increases with the evolutionary conservation of the site that undergoes nonsynonymous editing. To verify these predictions, we retrieved 573 protein-coding genes that have one-to-one orthologs in all 43 mammals examined (see Materials and Methods). We aligned these sequences and removed alignment gaps, and identified from the HEK293T data 32 synonymous and 53 nonsynonymous C-to-U editing sites (within TC motifs) in the human sequences of these alignments. Using the method based on Jensen–Shannon divergence (Capra and Singh 2007), we computed an amino acid conservation score for each codon that contains an edited site. As a control, we also identified unedited Cs that are in TC motifs in the human sequences of these alignments, and similarly computed conservation scores. We found no significant difference in conservation between codons with a synonymously edited C and codons with an unedited C that could potentially be synonymously edited (fig. 3A), suggesting that the occurrence of synonymous editing is unbiased with regard to the conservation of a codon. We similarly found no significant difference in conservation between codons with a nonsynonymously edited C and codons with an unedited C that could potentially be nonsynonymously edited (fig. 3B), suggesting that the occurrence of nonsynonymous editing is also unbiased with regard to the conservation of a codon. Hence, our findings provide no support to the adaptive hypothesis, although they may be attributable to a lack of statistical power.

The hypothesis that nonsynonymous editing is generally harmful further predicts a lower \( \frac{n}{s} \) ratio for genes that are under stronger evolutionary constraints. The evolutionary constraint on a gene is commonly measured by the ratio between the rate of nonsynonymous nucleotide substitution \( (d_{Ns}) \) and that of synonymous substitution \( (d_s) \). The lower the \( \frac{d_{Ns}}{d_s} \) ratio, the stronger the evolutionary constraint. In other words, we predict a positive correlation between \( \frac{n}{s} \) and \( \frac{d_{Ns}}{d_s} \). We computed the \( d_{Ns}/d_s \) ratio of each human gene by comparing with its one-to-one mouse ortholog. We then divided all genes into ten equal-sized bins based on the \( d_{Ns}/d_s \) ratio and computed the median \( d_{Ns}/d_s \) ratio of each bin.

Analyzing the HEK293T data set, we calculated the \( \frac{n}{s} \) ratio for each bin and found a significant positive correlation between \( \frac{n}{s} \) and median \( d_{Ns}/d_s \) across bins (Spearman’s correlation \( \rho = 0.70; P = 0.031; \) fig. 2B). We noticed that \( \frac{n}{s} \) > 1 for three bins with intermediate \( d_{Ns}/d_s \) (fig. 2B), but \( \frac{n}{s} \) does not significantly exceed 1 in any of these bins (nominal \( P = 0.402, 0.226, \) and 0.075, respectively).

We also analyzed the macrophage and monocyte data. Because the number of edited genes is much smaller in each of these data sets, we divided each data set into five bins. Although positive correlations were observed between \( \frac{n}{s} \) and median \( d_{Ns}/d_s \) across bins for both macrophages (\( \rho = 0.70, P = 0.23 \)) and monocytes (\( \rho = 0.69, P = 0.21 \)), they were not significant, possibly due to the small sample sizes (table 1).

Table 1. Data Sets Used in This Study.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Condition</th>
<th>Number of Edited Sites in CDS</th>
<th>Number of Edited Sites in TC Motifs (%)</th>
<th>Number of Edited Genes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macrophage</td>
<td>M1 polarization</td>
<td>93</td>
<td>89 (95.7)</td>
<td>85</td>
<td>Sharma et al. (2015)</td>
</tr>
<tr>
<td>Monocyte</td>
<td>Hypoxia</td>
<td>128</td>
<td>115 (89.8)</td>
<td>109</td>
<td>Sharma et al. (2015)</td>
</tr>
<tr>
<td>HEK293T</td>
<td>APOBEC3A overexpression</td>
<td>2,455</td>
<td>2,401 (97.8)</td>
<td>1,902</td>
<td>Sharma et al. (2017)</td>
</tr>
</tbody>
</table>

Fig. 1. Frequencies and editing levels of synonymous and nonsynonymous C-to-U RNA editing in human coding sequences. (A) Frequency of nonsynonymous editing is lower than that of synonymous editing in macrophages, monocytes, and HEK293T cells. P values are from \( \chi^2 \) tests. (B) Nonsynonymous editing levels are lower than synonymous editing levels. In the boxplot, the lower edge and upper edge of a box represent the 25% quartile \( (q_1) \) and 75% quartile \( (q_3) \), respectively. The horizontal line inside a box indicates the median \( (md) \). The whiskers extend to the most extreme values inside inner fences, \( md \pm 1.5(q_3-q_1) \). P values are from two-tailed Wilcoxon rank sum tests.

\( f_{n/s} \) from essential genes (fig. 2A). Hence, \( f_{n/s} \) is significantly lower for essential genes than nonessential genes \( (P = 0.003) \).

The hypothesis that nonsynonymous editing is generally harmful further predicts a lower \( f_{n/s} \) ratio for genes that are under stronger evolutionary constraints. The evolutionary constraint on a gene is commonly measured by the ratio between the rate of nonsynonymous nucleotide substitution \( (d_{Ns}) \) and that of synonymous substitution \( (d_s) \). The lower the \( d_{Ns}/d_s \) ratio, the stronger the evolutionary constraint. In other words, we predict a positive correlation between \( f_{n/s} \) and \( d_{Ns}/d_s \). We computed the \( d_{Ns}/d_s \) ratio of each human gene by comparing with its one-to-one mouse ortholog. We then divided all genes into ten equal-sized bins based on the \( d_{Ns}/d_s \) ratio and computed the median \( d_{Ns}/d_s \) ratio of each bin. Analyzing the HEK293T data set, we calculated the \( f_{n/s} \) ratio for each bin and found a significant positive correlation between \( f_{n/s} \) and median \( d_{Ns}/d_s \) across bins (Spearman’s correlation \( \rho = 0.70; P = 0.031; \) fig. 2B). We noticed that \( f_{n/s} > 1 \) for three bins with intermediate \( d_{Ns}/d_s \) (fig. 2B), but \( f_{n/s} \) does not significantly exceed 1 in any of these bins (nominal \( P = 0.402, 0.226, \) and 0.075, respectively).
Next, we correlated the conservation score and editing level for codons undergoing synonymous editing, and observed no significant correlation ($\rho = 0.13$, $P = 0.48$; fig. 3C). By contrast, a significant negative correlation was observed among sites undergoing nonsynonymous editing ($\rho = -0.28$, $P = 0.039$; fig. 3D). Hence, conserved sites tend to have reduced levels of nonsynonymous but not synonymous editing, which is opposite to the prediction of the adaptive hypothesis, but is consistent with the hypothesis that most nonsynonymous editing is deleterious.

**Discussion**

In this work, we performed the first transcriptome-wide test of the adaptive hypothesis of C-to-U editing of mRNAs originating from animal nuclear genomes. We found the
frequency of nonsynonymous editing lower than that of synonymous editing and the editing levels of nonsynonymous editing lower than those of synonymous editing. These results contrast predictions of the adaptive hypothesis. Instead, they suggest that most C-to-U RNA editing of coding sequences is nonadaptive, either neutral or deleterious. Consistent with the nonadaptive hypothesis, nonsynonymous editing, relative to synonymous editing, is rarer in essential genes and genes under relatively strong evolutionary constraints than in nonessential genes and genes under relatively weak constraints. Furthermore, we found no evidence for the conjecture that nonsynonymous editing is enriched at evolutionarily conserved sites. Instead, nonsynonymous editing levels tend to be lowered at evolutionarily conserved sites.

Because C-to-U coding RNA editing in HEK293T was detected only upon the artificial overexpression of the editing enzyme APOBEC3A, why did we observe signals of purifying selection against editing in this cell line? One possibility is that RNA editing is stochastic such that some cells have more nonsynonymous editing than other cells in a population of HEK293T cells. If cells having more nonsynonymous editing grows slower than cells having less nonsynonymous editing, \( f_s/f_e < 1 \) may be observed. Alternatively and more likely, what we observed in HEK293T reflects natural selection acting in tissues/cells where APOBEC3A is naturally highly expressed. Under this scenario, genes frequently coexpressing with APOBEC3A across human tissues should have been subjected to the strongest selection against editing. Hence, we predict that these genes are underrepresented among those edited in HEK293T upon APOBEC3A overexpression. To verify this prediction, we downloaded gene expression data of 53 human tissues from GTEx (www.gtexportal.org/). Using the R package WGCNA (Langfelder and Horvath 2008), we found from the GTEx data that 123 genes are located in the same coexpression module as APOBEC3A. Interestingly, none of these 123 genes are edited in the HEK293T data. By contrast, among the 19, 353 genes that do not share the coexpression module with APOBEC3A, 1872 are edited. Hence, genes that frequently coexpress with APOBEC3A across human tissues tend not to be edited upon APOBEC3A overexpression \( (P = 0.0005, \chi^2 \text{ test}) \), further supporting that C-to-U RNA editing is selected against in humans.

The generally nonadaptive nature of C-to-U coding RNA editing in humans begs for an explanation of its existence and prevalence. It is notable that nearly all APOBEC family members appear to inhibit retroviruses by the deamination of cytosines in nascent retroviral cDNAs, which is a powerful mechanism of innate immunity (Harris and Liddament 2004). The mechanism by which APOBEC3G restricts retroviral infection is best understood (Harris et al. 2003; Lecossier et al. 2003; Mangeat et al. 2003; Zhang et al. 2003). Briefly, when retroviruses enter host cells and generate the minus strand of DNA from reverse transcription of their RNA genomes, APOBEC3G converts Cs to Us in the DNA. Us are not supposed to exist in DNA and thus are excised by host DNA repair enzymes. However, these Us cannot be completely removed, resulting in mutations in newly synthesized viral RNA genomes that will reduce the viral fitness (Harris and Liddament 2004). This consideration suggests the possibility that the natural substrate of APOBEC3G is viral cDNA and that C-to-U RNA editing originates from its off-target activity (Sharma et al. 2016). Similarly, APOBEC3A is implicated in the inhibition of retroviruses, including human immunodeficiency virus HIV-1, human T-lymphotropic virus HTLV-1, human papillomavirus HPV, parvoviruses, and hepatitis B (Chen et al. 2006; Narvaiza et al. 2009; Berger et al. 2011; Sharma et al. 2017); it is likely that the viral cDNA is the natural substrate of APOBEC3A. This hypothesis is consistent with the high expression of APOBEC3A in macrophages and monocytes, which play important immune functions such as inhibiting HIV (Koppensteiner et al. 2012). In other words, the observed C-to-U RNA editing probably results from promiscuous targeting by APOBECs and was at least initially a deleterious byproduct of the beneficial functions of their antiviral activities. Note that, in the presence of the C-to-U RNA editing activity, a genomic site in a coding region that otherwise accepts T only can now accept C, because C can be converted to U at the RNA level. This may lead to the fixation of T-to-C mutations at a number of such sites. As a result, the RNA editing activity can no longer be removed, because it is impossible to reverse simultaneously all these Cs to Ts in the genome through evolution. Although there is little evidence that C-to-U RNA editing is beneficial, the above scenario explains why C-to-U editing is evolutionarily maintained (Covello and Gray 1993) and suggests the possible existence of sites where the editing is required for normal cellular function. But such sites, even if they are found, should not be automatically regarded as adaptations, because, at least in the above model, the present-day fitness is not higher than that before the origin of C-to-U RNA editing. *Bona fide* adaptive C-to-U RNA editing, reserved for beneficial tissue- or time-specific regulation of protein function, may be a minority among the already small proportion of editing that potentially plays physiological roles.

Our results on human C-to-U RNA editing are highly similar to previous findings of human A-to-I coding RNA editing (Xu and Zhang 2014) in that both are generally nonadaptive and likely reflect the off-target activities of respective editing enzymes. In addition, we discovered recently that most events of mA modification (methylation of the adenosine base at the nitrogen-6 position), a highly abundant internal modification of mRNA molecules, are nonadaptive (Liu and Zhang 2018). These similarities suggest a general trend of noisy posttranscriptional modification that requires us to rethink the functionality of such modifications. Specifically, it may be futile to assume that most events of any posttranscriptional modification are functional and search for their functions. Rather, searching for the function of a modification event only when there is indication that the modification is beneficial or evolutionarily conserved (Xu and Zhang 2015; Liu and Zhang 2018) may be a more productive approach.

Note, however, that our analysis focused on C-to-U RNA editing in coding sequences, whereas such editing is also known to occur in noncoding sequences, especially in 3′-untranslated regions (3′-UTR) (Blanc et al. 2014; Harjanto et al. 2016). It is possible that RNA editing at microRNA target sites in 3′-UTRs affects mRNA stability (Gu et al. 2012; Zhang
et al. 2017), although direct evidence is currently lacking. In the future, it will be interesting to investigate the biological significance of C-to-U RNA editing in noncoding regions. Furthermore, because the C-to-U RNA editing investigated in our study is primarily due to the activity of APOBEC3A, it will be important to verify if our findings apply to editing by other cytidine deaminases.

**Materials and Methods**

C-to-U editing sites in human coding regions were obtained from three publicly available data sets (Sharma et al. 2015, 2017). Specifically, edited sites in the first data set were identified by comparing whole transcriptomes of M1 and M2 macrophages that were generated in vitro from peripheral blood monocytes (Beyer et al. 2012; Sharma et al. 2015). Among different macrophage activation phases induced by IFNγ, C-to-U RNA editing is most prevalent in M1 and is significantly lowered in M2 (Sharma et al. 2015). A total of 93 sites in coding regions with a minimal editing level of 5% were considered to be edited. Edited sites in the second data set were identified from sequenced transcriptomes of monocytes under hypoxia (Sharma et al. 2015), and 128 sites in coding regions with a minimal editing level of 5% were considered to be edited. The third data set was derived from the transcriptome of HEK293T cells with APOBEC3A overexpressed, and 2,455 sites with a minimal editing level of 15% were considered to be edited (Sharma et al. 2017). All three data sets provided the chromosomal location, strand, flanking sequences, synonymous or nonsynonymous annotation, and editing level for each edited site. A total of 22,013 human protein coding genes were downloaded from Ensembl (GRCh38.10; Ensembl Genes 90). If a gene has several isoforms, only the coding sequence corresponding to the longest isoform was included in our analyses.

From each coding sequence, we first isolated all TC motifs using in-house Perl scripts and then examined the number of potential synonymous editing sites, which would lead to a synonymous change if edited from C to U, and the number of potential nonsynonymous editing sites, which would cause a nonsynonymous change if edited from C to U. All dS/dN values between human and mouse orthologous genes were retrieved from Ensembl using BioMart (http://www.ensembl.org/biomart/martview/; last accessed January 30, 2018). A total of 183 essential and 14,388 nonessential genes from humans were obtained from the OGEE database (Chen et al. 2017). We discarded conditional essential genes, which are essential in some but not all conditions. To calculate the conservation score of each amino acid position, we retrieved 574 aligned one-to-one orthologous genes of 43 mammals from the database OrthoMaM (Douzery et al. 2014). We calculated the evolutionary conservation score of each amino acid position using the method of Jensen–Shannon divergence (Capra and Singh 2007).

**Acknowledgments**

We thank members of the Zhang lab and two anonymous reviewers for valuable comments. This work was supported in part by a research grant from the U.S. National Institutes of Health (R01GM120093) to J.Z. Z.L. was supported by China Scholarship Council (201604910442).

**References**


C-to-U RNA Editing Is Nonadaptive


