Journal of Applied Genetics (2022) 63:413–421
https://doi.org/10.1007/s13353-022-00687-y

LETTER TO THE EDITOR

Poor evidence for host-dependent regular RNA editing in the transcriptome of SARS-CoV-2

Jinbao Zong[1] · Yanping Zhang[2] · Fenfen Guo[1] · Changyuan Wang[3] · Hong Li[1] · Gaoyang Lin[4] · Wenqing Jiang[5] · Xiaoping Song[1] · Xiufang Zhang[1] · Feng Huang[1] · Min Qi[1] · Xin Zheng[6]

Received: 20 October 2021 / Revised: 21 November 2021 / Accepted: 13 February 2022 / Published online: 18 February 2022

Introduction

Expression patterns of editing enzymes

Two deaminase protein families mediate the two types of RNA editing. A-to-I RNA editing is catalyzed by ADAR (adenosine deaminase acting on RNA) protein family (Savva et al. 2012), while C-to-U editing is mediated by APOBECs (Harris and Dudley 2015). Not all of these editing enzymes have access to the infected viruses. Three ADARs are encoded in mammalian cells. ADAR3, ADAR2, and the p110 isoform of ADAR1 are strictly located in nucleus, and only p150 isoform of ADAR1 could enter cytosol (Eisenberg and Levanon 2018; Patterson and Samuel 1995). APOBECs have a wider range of expression in both nucleus and cytosol but mainly nucleus (Safran et al. 2010). Under this scenario, APOBECs are more likely to interact with the viruses like SARS-CoV-2 to exert C-to-U editing, while only p150 isoform of ADAR1 has the chance to meet the virus. Moreover, ADARs are highly specifically expressed in nervous systems, while APOBECs have (relatively) less tissue-specificity (Harris and Dudley 2015; Safran et al. 2010; Tan et al. 2017). This limitation again hints that when SARS-CoV-2 invades the lungs, the dominant modification should be C-to-U editing rather than A-to-I editing (Harris and Dudley 2015; Li et al. 2020).

Regular editing and hyper-editing

Regarding both A-to-I and C-to-U RNA editing, there are different bioinformatic pipelines to identify these sites from the sequencing reads. Regular editing sites are identified by regular mapping and variant calling pipelines (Ramaswami et al. 2013). Under another situation, where editing sites are so abundant that the reads could not be aligned to the reference genome, one needs to transform the genome and re-map the reads, a methodology termed the hyper-editing pipeline (Porath et al. 2014). For SARS-CoV-2, in theory both regular editing and hyper-editing exist. Unless otherwise stated, the generally speaking editing sites refer to the regular editing sites (Di Giorgio et al. 2020). However, the regular editing sites are usually challenging to be accurately identified due to the genomic polymorphisms (SNPs) (Ramaswami et al. 2013). In contrast, hyper-editing pipeline requires the clustering of the same type of mismatch so that the randomly distributed SNPs are largely depleted (Porath et al. 2014). Therefore, when the identification of regular editing sites does not work well, investigating the hyper-editing sites would be a good choice.

Communicated by: Agnieszka Szalewska-Palasz

One sentence summary: Multiple aspects of evidence showed that the regular editing sites previously identified in SARS-CoV-2 were unreliable.

✉ Xin Zheng
zhengxin_qd@163.com

1 Department of Clinical Laboratory, The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong, China

2 Department of Anesthesiology, The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong, China

3 Department of Dermatology, The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong, China

4 Department of Thoracic Surgery, The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong, China

5 Department of Respiratory Diseases, The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong, China

6 Department of Respiratory and Critical Care Medicine, The Affiliated Qingdao Hiser Hospital of Qingdao University, Qingdao, Shandong, China

Variations in deaminase-knock-out cell lines were certainly not RNA editing sites

When looking for RNA editing sites, virtually no bioinformatic pipelines could completely get rid of the SNPs from RNA-sequencing data alone. The multiple filtering steps are only to reduce the false positive sites but are unable to completely erase the noises (Ramaswami et al. 2013). The following options may facilitate the accurate identification of RNA editing sites.

(1) DNA resequencing of the same sample to remove the DNA polymorphisms.
(2) Sanger sequencing of both DNA and RNA to see the RNA–DNA difference (RDD). However, Sanger sequencing is low throughput and could not verify thousands and millions of sites.
(3) When DNA resequencing and Sanger sequencing are not available, one could alternatively sequence the RNA of the ADAR-knock-out strains (for the A-to-I editing studies). Without editing by ADAR, the variations found in the RNAs could only come from the DNA polymorphisms.

This is also why multiple studies on A-to-I editing have used ADAR-knock-out strains to exclude the potential SNPs (Porath et al. 2014; Yu et al. 2016), and sometimes the only purpose of using ADAR-knock-out strain is to verify that the editing sites found in wild type are reliable (Porath et al. 2014). Regarding SARS-CoV-2, a positive single-strand RNA virus, it does not literally have a real genome. Therefore, using DNA resequencing or the RDD method to remove the SNPs is not applicable. The only way to increase the reliability of RNA editing sites is to use the ADAR or APOBEC knock-out lines. The variations in both wild-type and knock-out lines are genomic polymorphisms, and the variations specific to wild type should be RNA editing sites.

Intra-molecular and inter-molecular double-strand structures

APOBECs bind single-stranded RNAs (Harris and Dudley 2015; Salter and Smith 2018), while ADARs bind double-stranded RNAs (dsRNAs) (O'Connell et al. 2015; Quin et al. 2021). The dsRNAs could either be folded by a single RNA molecular, termed the intra-molecular hairpin, or formed by two reverse complementary RNAs, termed the inter-molecular dsRNA. Usually, the intra-molecular dsRNA is likely to be formed by inverted repeats, while the inter-molecular dsRNA could come from (1) the bi-directional transcription in eukaryote genomes or (2) the replication of ssRNA viruses (like SARS-CoV-2). Although SARS-CoV-2 has the potential to form inter-molecular dsRNA structure upon replication, there is still the possibility that SARS-CoV-2 could form intra-molecular hairpins. The dsRNAs from both sources could be targeted by ADAR. Any structural analyses related to RNA editing should seriously clarify which kind of dsRNA they refer to.

Aims and scopes

In this article, we aim to test the hypothesis presented by Di Giorgio et al (2020) and see whether we could fully reproduce their results and conclusions. The success of Di Giorgio et al. largely relies on the fact that SARS-CoV-2 is an RNA virus and the RNA editing could not be verified by the "traditional method in eukaryotes" (Sanger sequencing of both DNA and RNA to check the RNA–DNA difference). Therefore, the "untestability" of the sites give them confidence because no evidence could directly disprove these editing sites. However, this is also the limitation of their study because on the other hand no approaches could prove the authenticity of the RNA editing sites. The only plausible way to verify the RNA editing sites in an RNA virus is to construct an ADAR-knock-out line (host) as previous literatures performed (Porath et al. 2014). We provide several lines of evidence to show that the so-called RNA editing sites found by Di Giorgio et al. lack biological significance. Please also note that we only propose that those editing sites are unreliable (likely to be replication errors) but we do not claim that none of the sites they found are true editing sites. It is just a balance between signal and noise. True editing sites may be included in the total sites but the global candidate list is certainly a mixture of true sites and plenty of false positive sites.

Generally, proving something to be true is difficult but proving something to be false is easy. Proving all the A > G sites are A-to-I editing is difficult (as one should provide an extremely high A > G%), while proving not all the A > G sites are A-to-I editing is easy (as long as the A > G is not dominant). Therefore, our logic and conclusions are highly robust.

Materials and methods

The materials (RNA-sequencing datasets) and methods (bioinformatic pipeline) completely followed the Di Giorgio et al (2020) study, and no novel data and pipelines were added. The reference sequence of SARS-CoV-2 was downloaded from GISAID (https://www.gisaid.org/). We selected two transcriptome samples SRR10903401 and SRR10903402 to

repeat their analyses. Data were downloaded using command "prefetch" in Linux. Reads were aligned using BWA (Li and Durbin 2009). Variants were called using REDItools (Picardi and Pesole 2013) and JACUSA (Piechotta et al. 2017) after sorted by Samtools (Li et al. 2009). The minimum mapping quality (-q 25) and base quality (-Q 35) were retrieved from the command line provided in Di Giorgio et al. paper. The statistical tests were performed in the R platform.

Results and discussion

Lack of a clear prediction of the mismatch profile before analysis

Usually, studies on A-to-I RNA editing would have a very clear prediction of the final mismatch profile; that is, after multiple steps of filtering, the A > G should (or must) be the dominant type of mismatch (ideally > 80% should be A > G) (Li et al. 2014; Liscovitch-Brauer et al. 2017; Porath et al. 2014; Ramaswami et al. 2013). Even if they did not necessarily speak out this purpose of increasing the A > G fraction, their analysis would undoubtedly prove that the filtering steps and cutoffs were used in order to increase the A > G percentage. In contrast, Di Giorgio et al (2020) did not claim any expected pattern of the final mutation profile before the analysis, which sounds like "as long as we performed a series of filters, then whatever we found should be RNA editing sites." This take-for-granted logic was non-scientific and did not appear in any previous literatures on RNA editing.

Normally, when a basic anticipation of mutation profile was made, then one should apply a series of filtering steps to "enrich" the desired type of mismatch. In the following section, we would discuss that although the application of filters and cutoffs is good, it could be meaningless sometimes unless it remarkably changes the mutation profile or increases the A > G percentage.

Mismatches were not significantly enriched in both ends of reads

In the RNA editing studies of various species, it was commonly believed that the mismatches introduced by sequencing errors or misalignments were enriched at both ends of reads (Lin et al. 2012) so that researchers usually discarded several nucleotides at read ends to reduce the noises (Li et al. 2014; Ramaswami et al. 2013). Notably, if the signal to noise ratio was not elevated by removing the reads ends, then this step would be meaningless, because in that case, the trimming step would only (largely) reduce the number of mismatch sites while not reducing the false positive rate.

The Di Giorgio study indeed removed the 15 bp at read ends (Di Giorgio et al. 2020). However, when we repeated the two samples SRR10903401 and SRR10903402 with their pipeline, we did not find enrichment of mismatches at read ends because the non-mismatch sites had exactly the same distribution on reads (Fig. 1A). A depletion of nucleotides at the 5-prime end reflected the low base qualities (q) at read end (which were discarded by the "-q 25" parameter) (Di Giorgio et al. 2020). It suggested that using the whole reads did not increase the noise (and trimming the reads may not reduce the noise). Expectedly, we found that the relative fractions of 12 mismatch types were almost identical before and after removing the 15 bp (Fig. 1B). This meant that trimming the 15 bp did not improve the signal-to-noise ratio at all. What was the point of this trimming step?

Given the fact that the authors did not have a clear prediction of mismatch profile before the analysis, this trimming step became understandable. This was again the take-for-granted logic, as long as the authors utilized a series of steps to filter the mismatches, they would claim that "whatever we found should be RNA editing sites." Actually, a more plausible interpretation of these mismatches should be other polymorphisms or artefacts rather than RNA editing.

Raising the cutoff on depth did not improve the A > G percentage at all

In the field of RNA editing, the case of the paper by Li et al (2011) has been extensively discussed in the literatures. The authors found 12 types of RNA–DNA differences (RDDs) in human samples and claimed that there were 12 types of RNA editing in human (Li et al. 2011). Soon after its publication, numerous groups repeated the results and found that most of the RDDs were artefacts (Kleinman and Majewski 2012; Lin et al. 2012; Pickrell et al. 2012). Moreover, even using more stringent criteria to filter the mismatches (RDDs), researchers still obtained an A > G percentage of 15% (Lin et al. 2012). Then they concluded that "This low fraction (15% A > G) probably indicates a high false-positive rate (85% to 15%) even in the remaining set of RDD sites" (Lin et al. 2012).

Di Giorgio et al (2020) applied many cutoffs such as requiring minimum coverage = 20. However, this step did not change the mismatch profile at all (Fig. 2). Since the cutoff did not change the relative faction of mismatch types (did not increase A > G percentage), this filtering step became futile and meaningless because it merely reduced the number of variation sites but nothing more. The same went for their requirement of "reads supporting the SNV ≥ 4" (Di Giorgio et al. 2020).

A successful RNA editing paper would find an extremely high A > G% (usually > 80%) after multiple steps of filters (Liscovitch-Brauer et al. 2017). In contrast, the Di Giorgio

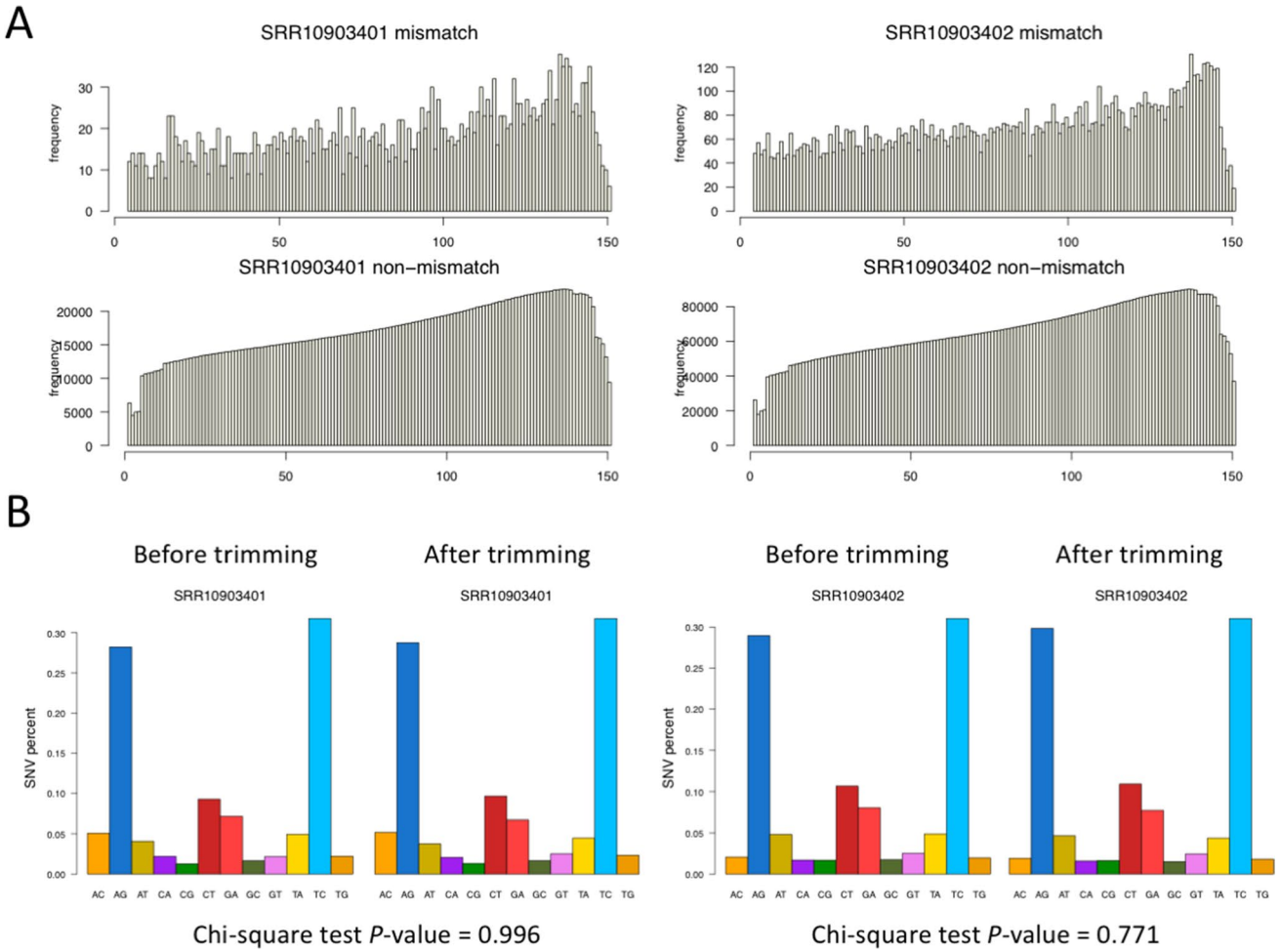

Fig. 1 The mismatches in two SARS-CoV-2 samples. **A** Position of mismatch or non-mismatch nucleotides on the reads. The nucleotides had base quality $q > 25$. **B** The fractions of 12 types of mismatches before and after trimming the 15 bp at both ends of reads. Chi-square tests were used to calculate the statistical significance

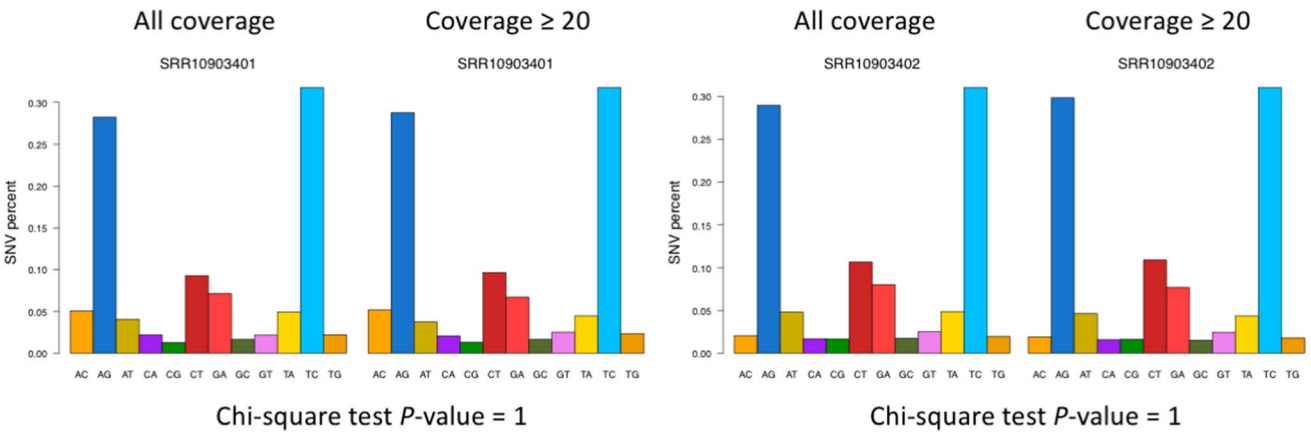

Fig. 2 The fractions of 12 types of mismatches before and after requiring coverage ≥ 20. Chi-square tests were used to calculate the statistical significance

et al (2020) paper was more similar to the Li et al (2011) paper, which:

(1) Did not have a prediction of what was expected to observe before the analysis, which indicates the take-for-granted logic that "as long as we performed a series of filters, then whatever we found should be RNA editing sites"
(2) Then obtained a mutation spectrum similar to the random mismatch spectrum, which was likely to be artefacts or SNPs

Briefly speaking, if one intended to call SNPs from the data, they would get the same mutations profile. Then how to prove that the mismatches found by Di Giorgio et al. were RNA editing sites? Interestingly, the authors said that "the SNV had a bias towards transitions (compared to transversion)" (Di Giorgio et al. 2020). When one talked about the terminology "transition and transversion," it should be noted that the natural mutations rate (replication errors, SNPs) would have a bias towards transition. This meant that the authors reduced the reliability of their own results by admitting a mutation bias towards transition.

Again, the more plausible interpretation of the mismatches should be replication errors (polymorphic sites) or sequencing errors (artefacts) rather than RNA editing. One could argue that requiring multiple reads supporting the same mutation site (say, SNV ≥ 4) would reduce the possibility of sequencing error. Nevertheless, the polymorphism assumption was still reasonable because these variation sites were subjected to numerous rounds of RNA replications and could absolutely have multiple reads supporting the same mutation site.

In a word, the golden standard for reliable RNA editing was to obtain a very high fraction of A > G (Liscovitch-Brauer et al. 2017), unless one used ADAR-knock-out lines to exclude the false positive mismatches (Porath et al. 2014; Yu et al. 2016).

Failure in interpreting the equally abundant T > C substitutions in this positive single-strand RNA virus

Di Giorgio et al (2020) obtained an equally high fraction of T > C mismatches compared to A > G mismatches. Normal readers would realize that this profile resembled the SNP profile and should not be regarded as RNA editing sites. Strikingly, Di Giorgio et al. just ignored the necessity to explain the T > C mismatches. They said "where A > G changes derive from deamination of A-to-I mediated by the ADARs. It is thus likely that the A > G/T > C changes seen in SARS-CoV-2 are also due to the action of ADARs" (Di Giorgio et al. 2020). How could one suddenly transit from A > G to A > G/T > C? After reading throughout the manuscript, readers could not figure out whether the authors regarded both A > G/T > C as A-to-I editing sites or just regarded A > G as the A-to-I editing sites. It was likely that the authors treated both A > G/T > C as editing sites (but without necessary explanation). Given their take-for-granted logic that "whatever we found should be RNA editing sites," it is not surprising that they were not willing to clarify this A > G/T > C issue.

In eukaryotes, the genes exist at both strands of the reference genome (because DNA is double-stranded). A-to-I RNA editing detection in eukaryotes would look for A > G mismatches on sense strand and T > C mismatches on antisense strand according to the variant calling format (Danecek et al. 2011). This did not mean that the T > C mutation took place on RNAs; it was just a format to record the A > G variations mapped to the antisense strand of reference genome. However, the SARS-CoV-2 genome was single sense strand; the A > I editing on RNA should directly create an A > G mismatch between the reads and the reference genome (Fig. 3). There was no reason to obtain many T > C mismatches, let alone regarding them as RNA editing. If an equally high fraction of T > C was observed, then one should be convinced that most of the variations should be polymorphisms. Di Giorgio et al (2020) may have underestimated the importance for giving a plausible explanation for the abundant T > C mismatches.

Intriguingly, in the "Fig. 4A and 4B" of their paper, Di Giorgio et al (2020) displayed the substitutions between different coronaviruses. The most abundant substitution type appeared to be C > T (clearly dominant and "asymmetric"), which was a strong indicator of prevalent C-to-U deamination by APOBEC. According to the logic of Di Giorgio et al., the mismatches derived from C-to-U deamination should also be "symmetric" just like the A > G/T > C caused by A-to-I deamination. However, Di Giorgio et al. did not give any explanations regarding why A-to-I editing leads to symmetric A > G and T > C mismatches but C-to-U editing only leads to a single asymmetrically dominant C > T peak.

Thus, all these evidences implied that a plausible interpretation of the A > G/T > C mismatches in the SARS-CoV-2 transcriptome data should be replication errors (polymorphisms) rather than RNA editing.

If editing in antisense strand existed, then the motif should be checked on antisense strand

One possible resource of the T > C mismatches was the RNA editing on antisense strand. This would make the virus genome similar to the eukaryote genome, where both strands of the reference genome could have RNAs (and thus could be edited). If this was the case, then the A-to-I RNA editing would:

(1) Create A > G mismatches on the reads mapped to sense strand

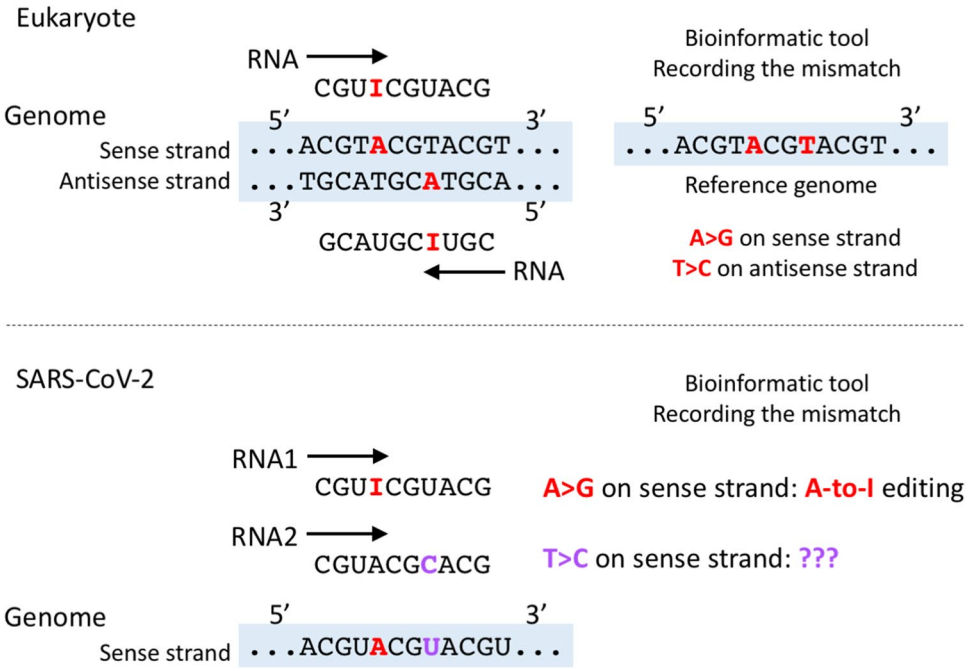

Fig. 3 The difference between eukaryote and SARS-CoV-2 genome. A-to-I editing in eukaryote mRNAs could lead to A > G on sense strand and T > C on antisense strand according to the variant-calling format. For SARS-CoV-2, A-to-I editing in RNA could only create A > G mismatches. The T > C mismatches mapped to the sense strand were meaningless

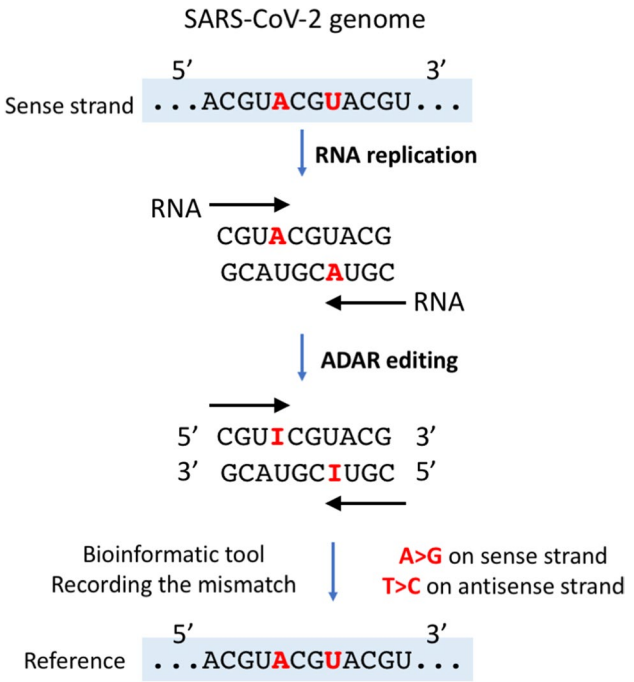

Fig. 4 The only possibility of T > C mismatch was the A-to-I editing on antisense RNA when SARS-CoV-2 was under RNA replication

(2) Create T > C mismatches on the reads mapped to the antisense strand according to the definition of variant-calling software (Fig. 4)

Another technical issue, the layout of the sequencing library (strand-specific or non-strand-specific), would also affect the interpretation on mutation sites. However, these details, together with the unexpected T > C sites, were not mentioned by Di Giorgio et al (2020). Mechanically applying the bioinformatic tools to a poorly understood dataset would only lead to a bunch of results with no biological significance.

Moreover, the ADAR preference motif was also problematic. As we said, readers could not figure out whether Di Giorgio et al. regarded both A > G/T > C as A-to-I editing sites or just regarded A > G as the A-to-I editing sites. Let us presume they use both A > G/T > C as A-to-I editing sites to look at the ADAR motif. Clearly, the − 1 and + 1 positions should be checked separately for A > G on sense strand and T > C on antisense strand (Fig. 5). Again, the authors did not clarify this at all (Di Giorgio et al. 2020). If one regarded both A > G/T > C as editing sites, and looked at their − 1 and + 1 positions at the sense strand of SARS-CoV-2 sequence, then the 3-mer motif would certainly be undesired. The part discussing the nucleotide context would be meaningless.

By observing the poor explanation on T > C variation and the erroneous analysis on ADAR motifs, it again indicated that a more plausible interpretation of the mismatches should be replication errors rather than RNA editing.

No association between editing sites and RNA structure was found

As introduced, ADAR binding could either take place on intra-molecular dsRNAs (usually hairpins) or the inter-molecular dsRNAs (when ssRNA viruses were

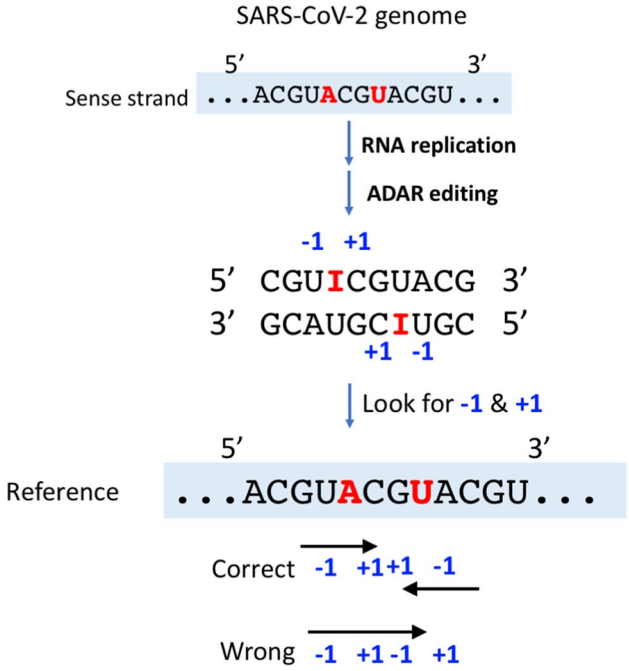

Fig. 5 How to recognize the − 1 and + 1 positions of the A-to-I editing sites in SARS-CoV-2 genome? The T > C mismatches (if regarded as antisense RNA editing) should be looked at in a reverse complementary direction. Otherwise, the − 1 and + 1 motif would be incorrect

replicating). The replication process was transient compared to the whole life cycle of SARS-CoV-2, there was no reason to solely mention ADAR binding to intermolecular dsRNAs (Di Giorgio et al. 2020). To be more objective, the hairpin formed by a single RNA molecular should be highlighted instead of the double strand formed by reverse complementary sequences during replication.

Besides, during SARS-CoV-2 replication, whether the RNA-dependent RNA polymerase (RDRP) would impede ADAR binding was never discussed (Di Giorgio et al. 2020). In theory, if the RDRP had a strong steric effect, then the antisense A-to-I editing (resulting in T > C mismatches when mapped to sense genome) should be very rare. In contrast, the hairpin formed by a single RNA molecule would be easily bound by ADAR. All these lines of evidence suggested that:

(1) The main point of the dsRNA topic should be intramolecular dsRNAs rather than inter-molecular dsRNA
(2) The T > C mismatches should be rare (due to RDRP) if they were derived from antisense editing

Expression pattern determined that APOBEC should play a more essential role than ADAR

It was established that only the p150 isoform of ADAR1 could enter the cytosol, while other ADARs were nucleus-located (Eisenberg and Levanon 2018; Patterson and Samuel 1995). The subcellular localization of APOBECs was much wider, including nucleus and cytosol (Safran et al. 2010) (although mainly nucleus). Therefore, one would expect that APOBECs were more likely to have access to the virus than ADARs. There was even no evidence to induce anyone to expect that A-to-I editing should be more prevalent than C-to-U editing in SARS-CoV-2.

Without further molecular evidence to prove that SARS-CoV-2 indeed frequently interacted with ADARs, the prevalent A-to-G found by Di Giorgio et al (2020) could only be the mutations introduced by replication errors (or even sequencing errors), which themselves have already admitted (more transitions than transversions). This further shed doubt on their interpretation of the origin of the variations. The only possible verification is to see whether these variations exist in ADAR-knock-out hosts. Again, Di Giorgio et al. largely benefited from the "untestability" of the so-called RNA editing sites, and this "untestability" allows them to confidently claim that "as long as we performed a series of filters, then whatever we found should be RNA editing sites."

No analysis on biological significance was performed to prove the authenticity of editing sites

Virtually all literatures studying the RNA editing in coding region would discuss whether the editing sites would change the amino acid (missense) or not (synonymous) (Alon et al. 2015; Jiang et al. 2021; Li et al. 2014; Liscovitch-Brauer et al. 2017; Liu and Zhang 2018; Xu and Zhang 2014; Yu et al. 2016, 2021; Zhang et al. 2021). While the synonymous editing sites were used as neutral control as they were not subjected to natural selection, the missense editing sites could either be over-represented if editing was adaptive (Liscovitch-Brauer et al. 2017) or be suppressed if editing was non-adaptive (Xu and Zhang 2014). There should be signals of natural selection hidden in the occurrence and frequency of editing sites. However, Di Giorgio et al. did not distinguish the different categories of editing sites (Di Giorgio et al. 2020). This was an indicator of unreliable editing sites if no selective patterns were found between missense and synonymous sites (Li et al. 2011). In theory, SARS-CoV-2 should be subjected to strong natural selection which forced the virus to adapt to the environment. The evolutionary and selective analyses on missense versus synonymous sites were simple, conceivable, and highly feasible without any technical limitations. There was no reason to omit these missense-synonymous analyses unless the outcomes were negative and unreported. Accordingly, using the data and pipeline provided by Di Giorgio et al. paper, we found that the observed occurrence of missense and synonymous

variation sites did not show difference to random expectation (p-value > 0.1), which implied that these variations were not subjected to natural selection and were likely to be artefacts.

Conclusions

All of our data and evidence showed that the so-called RNA editing sites found by Di Giorgio et al. were more likely to be polymorphisms in SARS-CoV-2 genome introduced by replication errors. We propose that the findings and conclusions made by Di Giorgio et al. might be flawed and misleading and should be corrected in instant.

Acknowledgements We thank all the colleagues that have given suggestions to this project.

Author contribution JZ, YZ, FG, CW, HL, GL, WJ, XS, XFZ, FH, and MQ analyzed the data and drafted this manuscript. XZ supervised this work and revised the manuscript. All authors including approved the submission and publication of this manuscript.

Data Availability The materials (RNA-sequencing datasets) and methods (bioinformatic pipeline) completely followed the Di Giorgio et al (2020) study, and no novel data and pipelines were added. The reference sequence of SARS-CoV-2 was downloaded from GISAID (https://www.gisaid.org/). We selected two transcriptome samples SRR10903401 and SRR10903402 to repeat their analyses.

Declarations

Ethics approval and consent to participate Not applicable. This study has used public data.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Alon S, Garrett SC, Levanon EY, Olson S, Graveley BR, Rosenthal JJ, Eisenberg E (2015) The majority of transcripts in the squid nervous system are extensively recoded by A-to-I RNA editing. Elife 4:e05198

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158

Di Giorgio S, Martignano F, Torcia MG, Mattiuz G, Conticello SG (2020) Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. Sci Adv 6:eabb5813

Eisenberg E, Levanon EY (2018) A-to-I RNA editing - immune protector and transcriptome diversifier. Nat Rev Genet 19:473–490

Harris RS, Dudley JP (2015) APOBECs and virus restriction. Virology 479–480:131–145

Jiang Y, Cao X, Wang H (2021) Mutation profiling of a limbless pig reveals genome-wide regulation of RNA processing related to bone development. J Appl Genet 62: 643–653. https://doi.org/10.1007/s13353-021-00653-0

Kleinman CL, Majewski J (2012) Comment on "Widespread RNA and DNA sequence differences in the human transcriptome." Sci 335:1302

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, Cheung VG (2011) Widespread RNA and DNA sequence differences in the human transcriptome. Sci 333:53–58

Li Q, Wang Z, Lian J, Schiott M, Jin L, Zhang P, Zhang Y, Nygaard S, Peng Z, Zhou Y et al (2014) Caste-specific RNA editomes in the leaf-cutting ant Acromyrmex echinatior. Nat Commun 5:4943

Li Y, Yang XN, Wang N, Wang HY, Yin B, Yang XP, Jiang WQ (2020) The divergence between SARS-CoV-2 and RaTG13 might be overestimated due to the extensive RNA modification. Future Virol 15:341–347

Lin W, Piskol R, Tan MH, Li JB (2012) Comment on "Widespread RNA and DNA sequence differences in the human transcriptome." Sci 335:1302

Liscovitch-Brauer N, Alon S, Porath HT, Elstein B, Unger R, Ziv T, Admon A, Levanon EY, Rosenthal JJC, Eisenberg E (2017) Trade-off between transcriptome plasticity and genome evolution in cephalopods. Cell 169(191–202):e111

Liu Z, Zhang J (2018) Human C-to-U coding RNA editing is largely nonadaptive. Mol Biol Evol 35:963–969

O'Connell MA, Mannion NM, Keegan LP (2015) The epitranscriptome and innate immunity. PLoS Genet 11:e1005687

Patterson JB, Samuel CE (1995) Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. Mol Cell Biol 15:5376–5388

Picardi E, Pesole G (2013) REDItools: high-throughput RNA editing detection made easy. Bioinformatics 29:1813–1814

Pickrell JK, Gilad Y, Pritchard JK (2012) Comment on "Widespread RNA and DNA sequence differences in the human transcriptome." Sci 335:1302

Piechotta M, Wyler E, Ohler U, Landthaler M, Dieterich C (2017) JACUSA: site-specific identification of RNA editing events from replicate sequencing data. BMC Bioinformatics 18:7

Porath HT, Carmi S, Levanon EY (2014) A genome-wide map of hyper-edited RNA reveals numerous new sites. Nature Communications 5:4726. https://doi.org/10.1038/ncomms5726

Quin J, Sedmik J, Vukic D, Khan A, Keegan LP, O'Connell MA (2021) ADAR RNA modifications, the epitranscriptome and innate immunity. Trends Biochem Sci 46:758–771

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. Nat Methods 10:128–132

Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H et al (2010) GeneCards Version 3: the human gene integrator. Database (Oxford) 2010:baq020

Salter JD, Smith HC (2018) Modeling the embrace of a mutator: APOBEC selection of nucleic acid ligands. Trends Biochem Sci 43:606–622

Savva YA, Rieder LE, Reenan RA (2012) The ADAR protein family. Genome Biol 13:252. http://genomebiology.com/2012/13/12/252

Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, Zhang R, Ramaswami G, Ariyoshi K et al (2017) Dynamic landscape and regulation of RNA editing in mammals. Nature 550:249–254

Xu G, Zhang J (2014) Human coding RNA editing is generally nonadaptive. Proc Natl Acad Sci U S A 111:3769–3774

Yu Y, Zhou H, Kong Y, Pan B, Chen L, Wang H, Hao P, Li X (2016) The landscape of A-to-I RNA editome is shaped by both positive and purifying selection. PLoS Genet 12:e1006191

Yu YY, Li Y, Dong Y, Wang XK, Li CX, Jiang WQ (2021) Natural selection on synonymous mutations in SARS-CoV-2 and the impact on estimating divergence time. Future Virol 16:447–450

Zhang YP, Jiang W, Li Y, Jin XJ, Yang XP, Zhang PR, Jiang WQ, Yin B (2021) Fast evolution of SARS-CoV-2 driven by deamination systems in hosts. Future Virol 16:587–590