SARS-CoV-2 evasion from ADAR hyper-editing is both genome-encoded and sustained by the virus replication strategy

Enrico Bortoletto  
University of Padua

Paola Venier  
University of Padua  https://orcid.org/0000-0001-9634-3662

Maxwell Shapiro  
Stony Brook University

Andrea Leonardi  
University of Padua  https://orcid.org/0000-0002-7246-8580

Umberto Rosani (✉ umberto.rosani@unipd.it)  
University of Padua

Brief Communication

Keywords: In-vivo Infection, RNA-seq Samples, Host dsRNAs, Dinucleotide Motifs, Coronovirus ORFs, Evolution Encoding

DOI: https://doi.org/10.21203/rs.3.rs-314516/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

We evaluated the role of ADAR during in-vivo SARS-CoV-2 infection, identifying ADAR-mediated hyper-editing only in 49 RNA-seq samples at a low level. Hyper-editing of host dsRNAs appeared not influenced by SARS-CoV-2 infection and showed higher efficiency compared to viral editing. Conversely, in mouse samples we found abundant hyper-editing with similar efficiency between host and SARS-CoV-2 RNAs. Underrepresentation of dinucleotide motifs along coronavirus ORFs suggested that SARS-CoV-2 resistance to ADAR hyper-editing is both evolutionary-encoded and sustained by viral replication strategy.

Introduction

Since 2000, Coronaviruses like SARS-CoV, MERS-CoV and SADS-CoV have caused pandemics in multiple hosts\textsuperscript{1}. Since the end of 2019, Severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) has rapidly spread worldwide, making COVID-19 the worst zoonotic disease of the modern history. Although considerably similar to coronaviruses of bats and pangolins\textsuperscript{2}, SARS-CoV-2 has jumped from a still unknown host into humans, with the cross-species transmission mediated by an efficient interaction with human cell receptors\textsuperscript{3}. Although the SARS-CoV-2 mutation rate is slower than other RNA viruses because of RNA proofreading activity\textsuperscript{4}, genomic data produced in real time have shown a strong bias towards C-to-T substitutions, possibly resulting from APOBEC-mediated editing\textsuperscript{5,6}.

SARS-CoV-2 mutations can crucially affect host-pathogen interactions by modifying the targets of RNA editing enzymes, the secondary structure of viral RNAs\textsuperscript{7} and stimulating innate immune responses via TLR7 and TLR8\textsuperscript{8}. RNA sequencing (RNA-seq) and biochemical analyses revealed an extraordinary activation of the immune system during SARS-CoV-2 infection, known as “cytokine storm”, which contributed to the SARS-CoV-2-induced mortality\textsuperscript{3}. Moreover, the footprint of APOBEC3 and ADAR deamination has been detected during SARS-CoV-2 infection \textit{in-vivo}\textsuperscript{6}. APOBEC and ADAR seem to deaminate in preferential contexts as APOBEC deaminates primarily T\textsubscript{C} or C\textsubscript{T} motifs in ssDNA, leading to TT or CT variations, whereas ADAR deaminates adenosine (A) to inosine (I) primarily in the context of WA (W=A/T) motifs in dsRNA, with the resulting inosine being read as guanine during translation. Depending on the host-virus combination, A-to-I hyper-editing \textit{(he)} can either limit the virus replication by impairing its dsRNAs and marking them for a rapid degradation\textsuperscript{9} or favors the virus by acting as negative-feedback in the immune system, reducing the dsRNA detection mediated by MDA5\textsuperscript{10}. The analysis of few SARS-CoV-2 RNA-seq samples outlined a low frequency of ADAR editing, and raises relevant questions about the ADAR role during infection\textsuperscript{11}.

In details: 1. To what extent is SARS-CoV-2 susceptible to ADAR-mediated editing? 2. Is ADAR activity, in the form of \textit{he}, traceable during SARS-CoV-2 infection? Here, we verified the \textit{in-vivo} levels of ADAR \textit{he} in 863 RNA-seq datasets from humans and other SARS-CoV-2 infection models. Analyzing the dinucleotide composition of the open-reading frames (ORFs) of SARS-CoV-2 and other coronaviruses, we traced the
evolutionary footprints of RNA editing enzymes and we estimated the current susceptibility of these coronaviruses to RNA editing.

Results And Discussion

We searched for SARS-CoV-2 transcription in 636 human-SARS-CoV-2 RNA-seq datasets, identifying 328 samples with at least 1,000 SARS-CoV-2 reads (S. Table 1). We applied the hyperediting tool\textsuperscript{12} on these human host samples to trace genuine ADAR hyper-editing (he), retrieving 14,273 SARS-CoV-2 he reads, with an average of 0.036 edited reads every thousand viral reads (S. Table 1). Notably, he along SARS-CoV-2 was identified only by lowering the fraction of minimum edited sites per read to 0.03 instead of 0.05. The he levels poorly correlated with the coverage of SARS-CoV-2 in these samples (Figure 1A, \(r=0.42\), p-values \(1.3 \times 10^{-8}\)) and only 49 samples included more than 100 he reads (for a total of 9,568 he reads). In these samples we showed that he is mostly localized in two conserved hotspots along the SARS-CoV-2 genome: one localized on the RNA-dependent RNA polymerase (nsp12, position 14221:14331) and the second on nsp6 (position 11058:11162), both encoded within the polycistronic ORF1ab (Figure 1D). Normalizing he by gene expression levels, we demonstrated that he mostly impacted ORF6 (Figure 1C). ORF6 is known to impair the transcriptional induction of ISGs by interacting with STAT1 and STAT2\textsuperscript{13}.

The analysis of additional 227 RNA-seq datasets referred to SARS-CoV-2 infection in non-human hosts, including hamster, ferret, non-human primates, and mice revealed relevant he levels exclusively in mouse samples of one experiment (PRJNA646535, S. Table 1). SARS-CoV-2 infection in these mice was made possible by ACE-transfection, whereas the role of Interferon signaling during infection was evaluated by knocking-out IFNR or IRF3/7\textsuperscript{14}. Notably, the IRF3/7 knock-out mice displayed the highest he levels (1.16‰), followed by control (0.26‰) and IFNR knock-out mice (0.11‰). In these mouse samples, we detected a higher efficiency of he, measured as the number of edited bases per he cluster, compared to SARS-CoV-2 in humans, and the normalized he levels homogenously impacted SARS-CoV-2 genes (Figure 1C and D).

To verify if SARS-CoV-2 could interfere with ADAR he of host dsRNAs, we traced he events on mouse and human genomes. In mice, ADAR he impacted preferentially non-coding genes, although he levels were not influenced by SARS-CoV-2 infection for either protein coding genes or non-coding elements (S. Figure 1A). Although ADAR expression levels appeared low (<3 TPMs) and mildly downregulated in knock-out mouse samples, we observed a reduced he levels in these latter samples, possibly due to the general reduction of interferon-related genes in which ADAR is included (S. Figure 1A). The efficiency of he was similar between SARS-CoV-2 and mouse in all the samples (Figure 1B). Unfortunately, human host reads have been removed from most of the samples with SARS-CoV-2 he. In 7 human samples we could show that the host he efficiency was higher compared to SARS-CoV-2, with an average of 5 edited bases per cluster compared to 3.5 (Figure 1B).
To further verify the possible implication of ADAR he during SARS-CoV-2 infection in humans, we analyzed post-mortem RNA-seq samples of lung with different SARS-CoV-2 infection levels (N=57). We demonstrated that he of host dsRNAs was not influenced by SARS-CoV-2 infection (S. Figure 1B). Similar to mouse samples, ADAR was only mildly modulated by SARS-CoV-2 (1.8x), with mid to low expression levels (<20 TPMs).

According to our results, we speculated that SARS-CoV-2 genome sequence could confer resistance to ADAR he. The analysis of the dinucleotide composition of ORFs belonging to 70 coronaviruses demonstrated a significant under-representation of the ‘WA’ (W=A/T), ‘TC’ and ‘CG’ motifs (S. Figure 2). While the ‘CG’ motif is known to be under-represented in coronaviruses, in agreement with the effect of the zinc finger antiviral protein (ZAP) and the low-CpG frequency characterizing vertebrate hosts, the ‘WA’ and ‘TC’ motifs are preferential targets for ADAR and APOBEC3 enzymes, respectively. No significant differences in the inter- and intra-genus under-representation were detectable for ‘WA’ and ‘TC’, except between alpha and delta-coronaviruses at ‘WA’ (p-value 0.0032, Figure 2, S. Table 2), with WA under-representation in Orf1ab and S detected in 96% and 81% of the tested coronavirus genomes, respectively. The second metric that we considered is “the replacement transition fraction”, or repTrFrac, which determines a significantly high mutation susceptibility in an ORF, leading to non-synonymous polymorphism (nsSNPs). We showed that repTrFrac in ‘WA’ was significantly higher than in TC for most of the coronavirus genomes (‘WA’: 78.8 ± 18%; ‘TC’: 31 ± 34.6%, p-value 2.79 e-11, Figure 2).

Among beta-coronaviruses, SARS-CoV-2 appears in the upper part of the distribution for ‘WA’ motifs, because of a significant WA under-representation in ORFs covering 95% of the genome (ORF1ab, S, ORF3a, M, ORF6, ORF7a, N), whereas ‘TC’ motifs SARS-CoV-2 are in the lower part of the distribution (ORF1ab and S, Figure 2). We also showed that the ‘WA’ under-representation characterizes most of the non-structural proteins encoded along Orf1ab, except for nsp7, nsp9-12 and nsp16 (Figure 1D).

**Conclusions**

We showed how ADAR and APOBEC have evolutionary contributed to minimize their own RNA editing targets in the extant coronavirus genomes. ADAR has massively directed genome evolution towards less editable targets, with just few spaces left for additional synonymous variations, whereas APOBEC has leaved room for additional synonymous variations. This result is consistent with the strong bias towards C-to-T variations traced during the SARS-CoV-2 pandemic, likely produced by APOBEC. Genome-encoded resistance could explain the low frequency of ADAR he on SARS-CoV-2 in-vivo, although the evidence of abundant he in mouse samples rather suggests that host-specific SARS-CoV-2 replication mechanisms contribute to reduce ADAR he. SARS-CoV-2 RNAs are protected in double membrane vesicles in humans, effectively masking the potential ADAR substrate for editing. In mice the formation of these vesicles has never been verified for SARS-CoV-2. Unmodified he levels towards human dsRNAs discharged the hypothesis of a direct interaction between SARS-CoV-2 and ADAR, according to the
absence of ADAR in the viral interactome\textsuperscript{20}. We concluded that SARS-CoV-2 escapes ADAR he because of a combination of genome-encoded resistance and protected virus replication mechanisms in humans.

**Materials And Methods**

**Data retrieving.** Genome sequences of 70 coronaviruses were downloaded from the NCBI genome database and parsed to extract 617 open reading frames (ORFs, S. Table 2). A total of 1,792 RNA-seq datasets of SARS-CoV-2 were retrieved form NCBI SRA archive (accessed 1\textsuperscript{st} of December 2020) and screened as follows: RNA-seq metadata were used to extract samples of in-vivo infection in different host, while the number of SARS-CoV-2 reads per sample was assessed by mapping the reads to the MN908947 reference genome using \textit{bwa} (github.com/lh3/bwa). Samples with at least 1,000 SARS-CoV-2 reads were considered positive.

**ADAR hyper-editing analysis.** The \textit{hyperediting} tool\textsuperscript{12} was applied after minimal modifications of the original version, which utilized \textit{bwa}, SAMtools (github.com/samtools) and BEDTools (github.com/arq5x/bedtools2), implemented to overcome software incompatibilities. The tool parameters were adapted to our model, applying: 3/5 for \textit{Minimum of edited sites at Ultra-Edit read (\%)}; 60 for \textit{Minimum fraction of edit sites/mismatched sites (\%)}; 25 for \textit{Minimum sequence quality for counting editing event (PHRED)}; 60 for \textit{Maximum fraction of same letter in cluster (\%)}; 20 \textit{Minimum of cluster length (\%)}; and imposing that the he clusters should not be completely included in the first or last 20\% of the read. Outputs in BED format were parsed using custom scripts, and further analyzed using CLC Genomic Workbench v.21 (Qiagen, US).

**Gene expression analysis and he normalization.** Quality-trimmed reads were mapped to the SARS-CoV-2 reference genome (MN908947) applying 0.8 and 0.8 for length and similarity parameters, respectively. Gene expression values were computed as Transcript Per Million (TPM) or as uniquely mapped reads, to normalize he levels.

**Under-representation analysis.** Under-representation and replacement transition fraction analysis were performed using the n3 module of the \textit{Cytidine Deaminase Representation Reporter} (CDUR)\textsuperscript{21}. Briefly, this reporter received as input a coding sequence, which was shuffled 1,000 times by switching nucleotides in the third positions of the codons while maintaining the integrity of the amino-acid sequence as well as the genome GC content. We measured the relevant statistics (e.g., \textit{belowTA} and \textit{repTrFracTA}) as follows. The “\textit{below}” metrics counted the number of hotspots (e.g., TA) in the input and compared this number to the distribution of hotspots observed in the shuffled sequences to obtain an empirical P-value. The replacement transition fraction, or \textit{repTrFrac}, compared the ratio of possible non-synonymous mutations that can occur at the hotspot (e.g., TA) to the observed number of hotspots, obtaining a P-value in a similar way. This fraction was compared to the distribution resulting from the shuffled sequences, to obtain a second empirical P-value.
Statistical analysis. Since the data were not normally distributed (Shapiro-Wilk test), we applied the Kruskal-Wallis test and the Mann–Whitney U test for analyzing the specific sample pairs for stochastic dominance. To examine the correlation between the number of viral and edited reads we calculated the Spearman’s rank correlation coefficient. All the statistical analyses were performed using R (version 4.0.3)\(^{22}\).

References
1. Zheng, J. SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. *Int. J. Biol. Sci.* **16**, 1678–1685 (2020).
2. Lam, T. T.-Y. *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).
3. Song, P., Li, W., Xie, J., Hou, Y. & You, C. Cytokine storm induced by SARS-CoV-2. *Clin. Chim. Acta Int. J. Clin. Chem.* **509**, 280–287 (2020).
4. Sevajol, M., Subissi, L., Decroly, E., Canard, B. & Imbert, I. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. *Virus Res.* **194**, 90–99 (2014).
5. Simmonds, P. Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. *mSphere* **5**, (2020).
6. Di Giorgio, S., Martignano, F., Torcia, M. G., Mattiu, G. & Conticello, S. G. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci. Adv.* **6**, eabb5813 (2020).
7. Simmonds, P. Pervasive RNA Secondary Structure in the Genomes of SARS-CoV-2 and Other Coronaviruses. *mBio* **11**, (2020).
8. Kosuge, M., Furusawa-Nishii, E., Ito, K., Saito, Y. & Ogasawara, K. Point mutation bias in SARS-CoV-2 variants results in increased ability to stimulate inflammatory responses. *Sci. Rep.* **10**, 17766 (2020).
9. Shevchenko, G. & Morris, K. V. All I’s on the RADAR: role of ADAR in gene regulation. *FEBS Lett.* (2018) doi:10.1002/1873-3468.13093.
10. Pestal, K. *et al.* Isoforms of the RNA editing enzyme ADAR1 independently control nucleic acid sensor MDA5-driven autoimmunity and multi-organ development. *Immunity* **43**, 933–944 (2015).
11. Picardi, E., Mansi, L. & Pesole, G. A-to-I RNA editing in SARS-COV-2: real or artifact? *bioRxiv* 2020.07.27.223172 (2020) doi:10.1101/2020.07.27.223172.
12. Porath, H. T., Carmi, S. & Levanon, E. Y. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.* **5**, 4726 (2014).
13. Miorin, L. *et al.* SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28344–28354 (2020).
14. Israelow, B. *et al.* Mouse model of SARS-CoV-2 reveals inflammatory role of type I interferon signaling. *J. Exp. Med.* **217**, (2020).
15. Desai, N. et al. Temporal and spatial heterogeneity of host response to SARS-CoV-2 pulmonary infection. *Nat. Commun.* **11**, 6319 (2020).

16. Nchioua, R. et al. SARS-CoV-2 Is Restricted by Zinc Finger Antiviral Protein despite Preadaptation to the Low-CpG Environment in Humans. *mBio* **11**, (2020).

17. Lamers, M. M., van den Hoogen, B. G. & Haagmans, B. L. ADAR1: ‘Editor-in-Chief’ of Cytoplasmic Innate Immunity. *Front. Immunol.* **10**, 1763 (2019).

18. Chen, J. & MacCarthy, T. The preferred nucleotide contexts of the AID/APOBEC cytidine deaminases have differential effects when mutating retrotransposon and virus sequences compared to host genes. *PLoS Comput. Biol.* **13**, (2017).

19. Wolff, G., Melia, C. E., Snijder, E. J. & Bárcena, M. Double-Membrane Vesicles as Platforms for Viral Replication. *Trends Microbiol.* **28**, 1022–1033 (2020).

20. Mourier, T. et al. Host-directed editing of the SARS-CoV-2 genome. *Biochem. Biophys. Res. Commun.* **538**, 35–39 (2021).

21. Shapiro, M., Meier, S. & MacCarthy, T. The cytidine deaminase under-representation reporter (CDUR) as a tool to study evolution of sequences under deaminase mutational pressure. *BMC Bioinformatics* **19**, 163 (2018).

22. R: The R Project for Statistical Computing. https://www.r-project.org/.

**Supplementary Material Legends**

**Supplementary Figure 1.** ADAR hyper-editing oh host dsRNAs. The number of hyper-edited genes in mouse (A, N=8) and human (B, N=57) samples is reported separately for coding (left) and non-coding (right) genes. The orange lines (secondary axes) indicate the average level of normalized hyper-editing in these samples. Mouse samples are grouped by condition: wt, wild type; ace, ace-transfected; ifnr-, INFR knock-out; irf3/7-, IRF3/7 knock-out. Human lung samples are grouped by SARS-CoV-2 infection level: high, more than 50 TPMs of SARS-COV-2 expression; low, 5-50 TPMs; no, less than 5 TPMs.

**Supplementary Figure 2.** Under-representation analysis of di-nucleotide motifs along coronavirus ORFs. In the boxplots, the analyzed dinucleotide motifs are reported along the x axis, and the y axis values refer to the average percentage of ORFs with an under-representation of the corresponding dinucleotide on the total number of ORFs for each virus. The * and *** symbols indicate p-value smaller than 0.05 and 0.001, respectively, for each of the possible comparisons between dinucleotide motifs (Mann–Whitney U test).

**Supplementary Table 1.** List of the selected RNA-seq datasets. The Bioproject and run IDs, library layout and selection method, organism and sequenced tissue are reported. Also, the number of sequencing reads, SARS-CoV-2 (viral) reads, ADAR hyper-edited reads and hyper-editing ratio on the virus are reported.

**Supplementary Table 2.** Under-representation analysis. For each coronavirus ORF, the position in the viral genome, the genus, NCBI ID and description of the virus are described. The below columns report the p-
value associated with the under-representation of the specified dinucleotide whereas the \( \text{repTrFrac} \) columns report the p-value associated with the replacement transition fraction of the specified dinucleotide.

**Declarations**

**Acknowledgments**

University of Padova Strategic Research Infrastructure Grant 2017: “CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l’Innovazione”. DOR 2020 granted to PV.

**Author contributions**

EB and UR designed the analysis; EB and UR analyzed hyper-editing data; EB and MS analyzed coronavirus genome data; UR analyzed RNA-seq data; EB, MS, AL and PV contributed to discussion and to the homogenization of the different analyses; UR wrote the manuscript; PV and EB revised and improved the manuscript.

**Competing Interests statement**

The authors declare no competitive interests.