CORONAVIRUS

Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2

Salvatore Di Giorgio[1,2]*, Filippo Martignano[1,2]*, Maria Gabriella Torcia[3], Giorgio Mattiuz[1,3†‡], Silvestro G. Conticello[1,4†‡]

The COVID-19 outbreak has become a global health risk, and understanding the response of the host to the SARS-CoV-2 virus will help to combat the disease. RNA editing by host deaminases is an innate restriction process to counter virus infection, but it is not yet known whether this process operates against coronaviruses. Here, we analyze RNA sequences from bronchoalveolar lavage fluids obtained from coronavirus-infected patients. We identify nucleotide changes that may be signatures of RNA editing: adenosine-to-inosine changes from ADAR deaminases and cytosine-to-uracil changes from APOBEC deaminases. Mutational analysis of genomes from different strains of Coronaviridae from human hosts reveals mutational patterns consistent with those observed in the transcriptomic data. However, the reduced ADAR signature in these data raises the possibility that ADARs might be more effective than APOBECs in restricting viral propagation. Our results thus suggest that both APOBECs and ADARs are involved in coronavirus genome editing, a process that may shape the fate of both virus and patient.

INTRODUCTION

Emerging viral infections represent a threat to global health, and the recent outbreak of novel coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, novel coronavirus, 2019-nCoV) exemplifies the risks (*1*, *2*). As viruses are obligate intracellular parasites, organisms have evolved innate immune mechanisms to sense and counter the viruses. Among these mechanisms, RNA and DNA editing mediated by endogenous deaminases can provide a potent defense against specific viruses. Two deaminase families are present in mammalian species: The ADARs (adenosine deaminases that act on RNA) target double-stranded RNA (dsRNA) for deamination of adenines into inosines (A-to-I) (*3*, *4*), and the APOBECs deaminate cytosines into uracils (C-to-U) on single-stranded nucleic acids [single-stranded DNA (ssDNA) and single-stranded RNA (ssRNA)] (*5*, *6*). During viral infections, ADARs act either directly, through hypermutation of the viral RNA, or indirectly, through editing of host transcripts that modulate the cellular response (*7*–*18*). On the other hand, APOBECs target the viral genome, typically DNA intermediates (*19*–*26*), either through C-to-U hypermutation or through a non-enzymatic path that interferes with reverse transcription (*27*, *28*). Although some APOBEC3 proteins can interfere in vitro with Coronaviridae replication, it is not clear whether their enzymatic activity is involved (*29*). Ultimately, though, these restriction systems can also be exploited by the viruses to support their infectivity and increase their evolutionary potential (*9*, *11*–*15*, *30*–*32*).

RESULTS

To assess whether RNA editing could be involved in human host responses to SARS-CoV-2 infections, we started from publicly available RNA sequencing datasets from bronchoalveolar lavage fluids (BALF) obtained from patients diagnosed with COVID-19. While transcriptomic data for all samples could be aligned to the SARS-CoV-2 reference genome, the quality of the sequencing varied and only eight samples had coverage and error rates suitable for the identification of potentially edited sites (data S1). We called single-nucleotide variants (SNVs) on these eight samples (*33*, *34*) using REDItools 2 (*35*–*37*) and JACUSA (*38*) using the following thresholds: reads supporting the SNV ≥4, allelic fraction ≥0.5%, coverage ≥20, quality of the reads >25, base quality >35 (fig. S1A). The two pipelines gave comparable results with ~50% of the SNV positions called by both (figs. S1B and S2). We identified 910 SNVs common to REDItools 2 and JACUSA, ranging from 24 to 238 SNVs per sample (Fig. 1 and data S3). Given the thresholds used to call the SNV, samples with lower sequencing depths displayed lower numbers of SNVs.

While the weight of each SNV type varies across samples (Fig. 1), a bias toward transitions is always present, which is even more evident when all mutational data are pooled (Fig. 2, A and B). This pattern holds true even when only SNVs recurring in more samples are considered (Fig. 2C).

The SNV frequency and number of transversions are compatible with the mutation rates observed in coronaviruses [$10^{-6/-7}$; (*39*)] and commonly associated to the RNA-dependent RNA polymerases (RdRps). RdRps are error prone and are considered the main source of mutations in RNA viruses. However, the coronavirus *nsp14-ExoN* gene provides a form of error correction (*40*), which is probably the reason mutation rates in coronaviruses are lower than those observed in RNA viruses with smaller genomes. The mutational spectrum in SARS quasispecies presents a very weak bias toward U-to-G. Inactivation of nsp14-ExoN error correction reveals the mutational spectrum of the RdRp, which is quite different from the pattern we observe (i.e., main changes are C-to-A, followed by U-to-C, G-to-U, A-to-C, and U-to-G) (*41*). Hence, we would consider that SNVs deriving from RdRp errors represent a marginal fraction of the SNVs in the SARS-CoV-2 samples.

The bias toward transitions—mainly A>G/T>C changes—resembles the pattern of SNVs observed in human transcriptomes (*42*) or in viruses (*8*, *10*, *18*), where A>G changes derive from deamination of

[1]Core Research Laboratory, ISPRO, Firenze 50139, Italy. [2]Department of Medical Biotechnologies, University of Siena, Siena 53100, Italy. [3]Department of Experimental and Clinical Medicine, University of Florence, Firenze 50139, Italy. [4]Institute of Clinical Physiology, National Research Council, 56124 Pisa, Italy.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: giorgio.mattiuz@unifi.it (G.M.); s.conticello@ispro.toscana.it (S.G.C.)

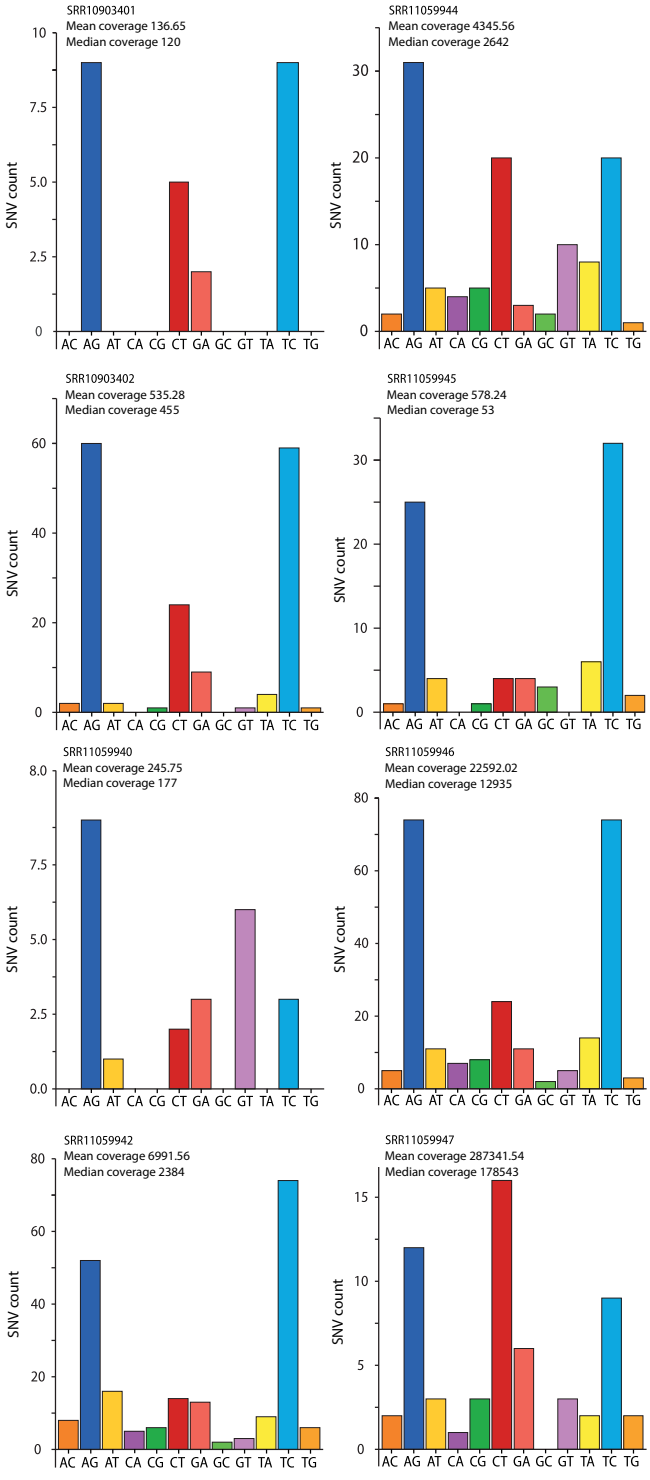

Fig. 1. SNVs identified in SARS-CoV-2 transcriptomes. The bar charts show the number of SNVs identified in each SARS-CoV-2 transcriptome for each SNV type (e.g., A>C, AC). The sequencing depth for each sample is indicated.

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.

C>T and G>A SNVs are the second main group of changes and could derive from APOBEC-mediated C-to-U deamination. Unlike A-to-I editing, C-to-U editing is a relatively rare phenomenon in the human transcriptome (*42*), and with regard to viruses, it has been associated only with positive-sense ssRNA rubella virus (*32*), where C>T changes represent the predominant SNV type. The observation that only A-to-I editing is present in RNA viruses that infect nonvertebrate animals, where RNA-targeting APOBECs are not present (*10*, *18*), supports the hypothesis that APOBECs are involved in the RNA editing of this human-targeting virus.

A third group of SNVs, A>T/T>A transversions, is also present in these samples. While this type of SNV has been reported in other genomic studies (*43*), its origin is still unknown.

A>G and T>C changes are evenly represented with respect to SNV frequency (Fig. 2A), the number of unique SNVs (Fig. 2, B and C), and their distribution across the viral genome (Fig. 2D). As ADARs target dsRNA, this suggests that dsRNA encompasses the entire genome. While dsRNA in human transcripts is often driven by inverted repeats, the most likely source of dsRNA in the viral transcripts is replication, where both positive and negative strands are present and can result in wide regions of dsRNA.

Unlike A-to-I changes, C-to-U changes are biased toward the positive-sense strand (Fig. 2, B to D; $P < 0.0001$). Because ADARs and APOBECs selectively target dsRNA and ssRNA, this distribution could arise from the presence at all times of RNA in a dynamic equilibrium between double-strandedness—when negative-sense RNA is being transcribed—and single-strandedness—when nascent RNA is released. Although some areas seem to bear fewer SNVs, these reduced SNV frequencies might be related to lower sequencing depth in those regions.

As APOBEC deaminases preferentially target cytosines within specific sequence contexts, we analyzed the nucleotide context of A-to-I and C-to-U SNVs in the viral genome (Fig. 3, A and B). A slight depletion of G bases in position −1 is present at A-to-I edited positions. This depletion is not as strong as the signal previously reported in human transcripts (*44*–*47*). The low editing frequencies we observe resembles the editing present on human transcripts containing Alu sequences, which were found in a limited number in those early datasets. There is no evidence of a sequence context preference if we use a larger dataset such as REDIportal (*48*), which includes >1.5 M sites in Alu repeats (fig. S3).

On the other hand, C-to-U changes preferentially occur downstream from uridines and adenosines, within a sequence context that resembles the one observed for APOBEC1-mediated deamination ([AU]C[AU]) (*49*, *50*).

We then aligned available genomes from SARS-CoV-2, Middle-East respiratory syndrome–related coronavirus (MERS-CoV), and SARS-CoV to test whether RNA editing could be responsible for some of the mutations acquired through evolution. The genomic alignments reveal that a substantial fraction of the mutations in all strains could derive from enzymatic deaminations (Fig. 4, A to C), with a prevalence of C-to-U mutations, and a sequence context compatible with APOBEC-mediated editing also exists in the genomic C-to-U SNVs (Fig. 4, D to F).

DISCUSSION

Our data source—metagenomic sequencing—raises the question whether the low-level editing we observe (~1%) reflects the actual levels of editing of viral transcripts within human cells. Aside from a small fraction of cellular transcripts edited at high frequency, most

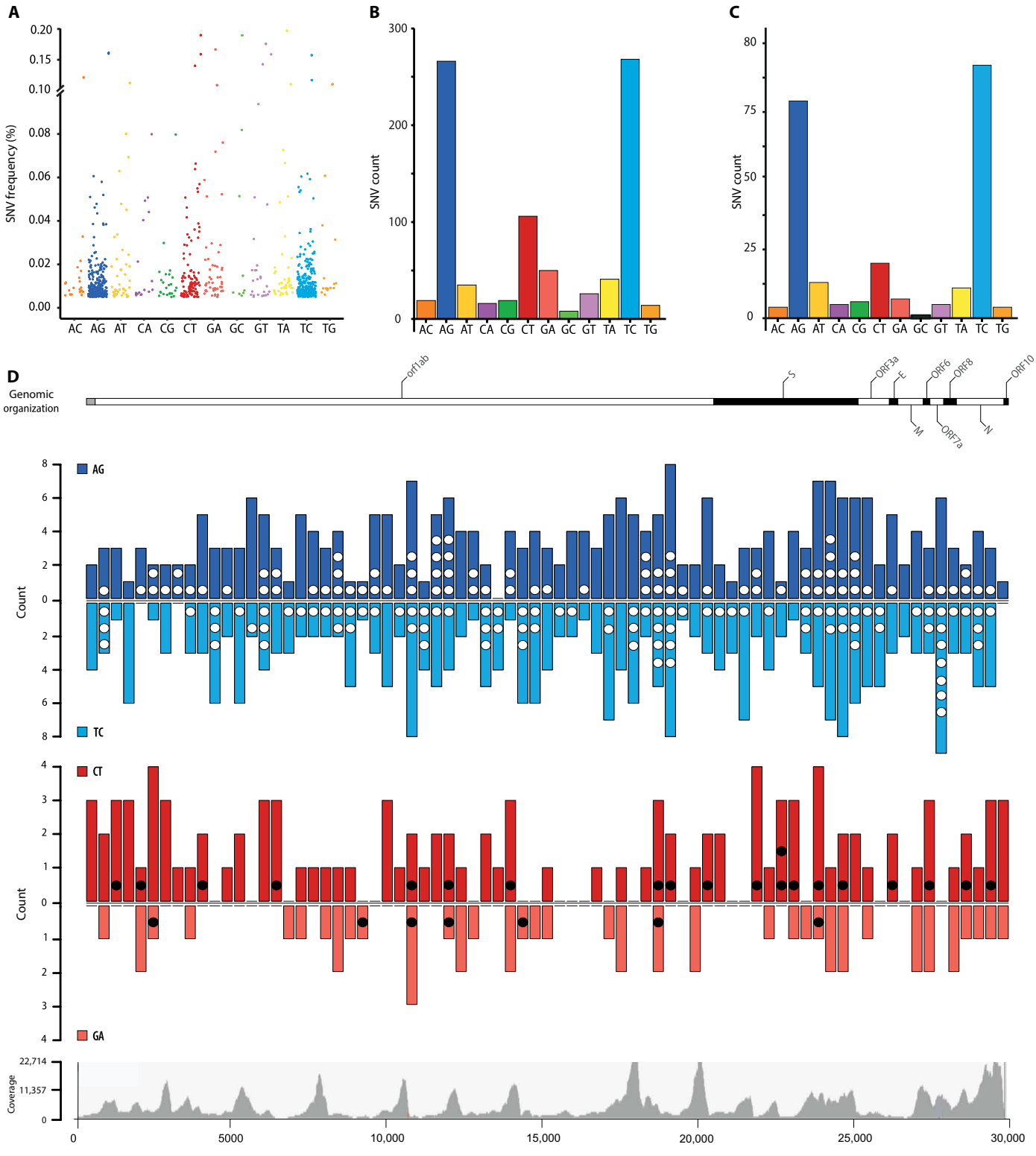

Fig. 2. SNV identified in SARS-CoV-2 transcriptomes. (**A**) Allelic fraction and (**B**) number of SNVs for each nucleotide change in the entire dataset and (**C**) for SNVs recurring in at least two samples. (**D**) Distribution of SNVs across the SARS-CoV-2 genome. A-to-G (blue) and C-to-U (red) SNVs are grouped in 400-nucleotide (nt) bins and plotted above (AG and CT) or below the line (TC and GA) based on the edited strand. Dots (white/black) indicate recurring SNVs. Genetic organization of SARS-CoV-2 (top). The dark/white shading indicates the viral coding sequences; coverage distribution of all analyzed samples (bottom).

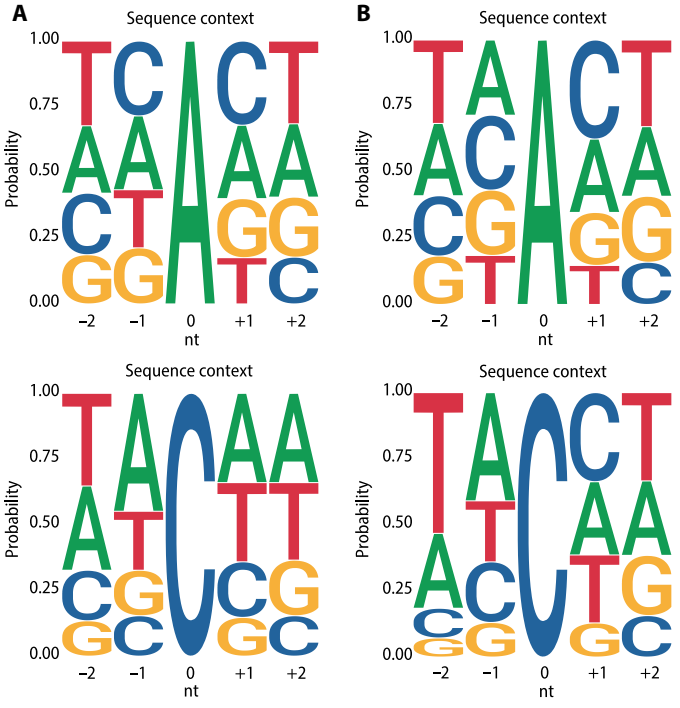

Fig. 3. Sequence contexts for SARS-CoV-2 RNA edited sites. (**A**) Local sequence context for A-to-I and C-to-U edited sites in the viral transcriptome and (**B**) for recurring sites.

ADAR-edited sites in the human transcriptome (typically inside Alu sequences) present editing levels of ~1% (*4*, *42*, *51*). It has been shown that a fraction of the cellular transcripts are hyperedited by ADARs (*52*–*54*). While we were unable to observe hyperedited reads in the metagenomic samples, it is possible that hyperedited transcripts fail to be packaged into the virus.

With regard to APOBEC-mediated RNA editing, its detection in the viral transcriptomes is already indicative, as this type of editing is almost undetectable in human tissues (*42*). This enrichment points either toward an induction of APOBECs triggered by coronavirus infection or to specific targeting of the APOBECs to the viral transcripts. APOBECs have been proved effective against many viral species in experimental conditions, yet, until now, their mutational activity in clinical settings has been shown only in a handful of viral infections (*19*–*26*) through DNA editing and, in rubella virus, on RNA (*32*).

As in rubella virus, we observe a bias in APOBEC editing toward the positive-sense strand. This bias and the low editing frequencies might be indicative of the dynamics of the virus, from transcription to selection of viable genomes. It is reasonable to assume that sites edited on the negative-sense strand will result in a mid-level editing frequency, as not all negative-sense transcripts will be edited (Fig. 5A). On the other hand, editing of the positive-sense strand can occur upon entry of the viral genome, thus yielding high-frequency editing (Fig. 5B), or after viral genome replication, resulting in low-frequency editing (Fig. 5C). The lack of a sizable fraction of highly edited C>T SNVs suggests that APOBEC editing occurs late in the viral life cycle (Fig. 5C). Yet, because they occur earlier, G>A SNVs should be closer in number to C>T SNVs and with higher levels of editing, which is not what we observe (Fig. 2, A to C). The overrepresentation of C>T

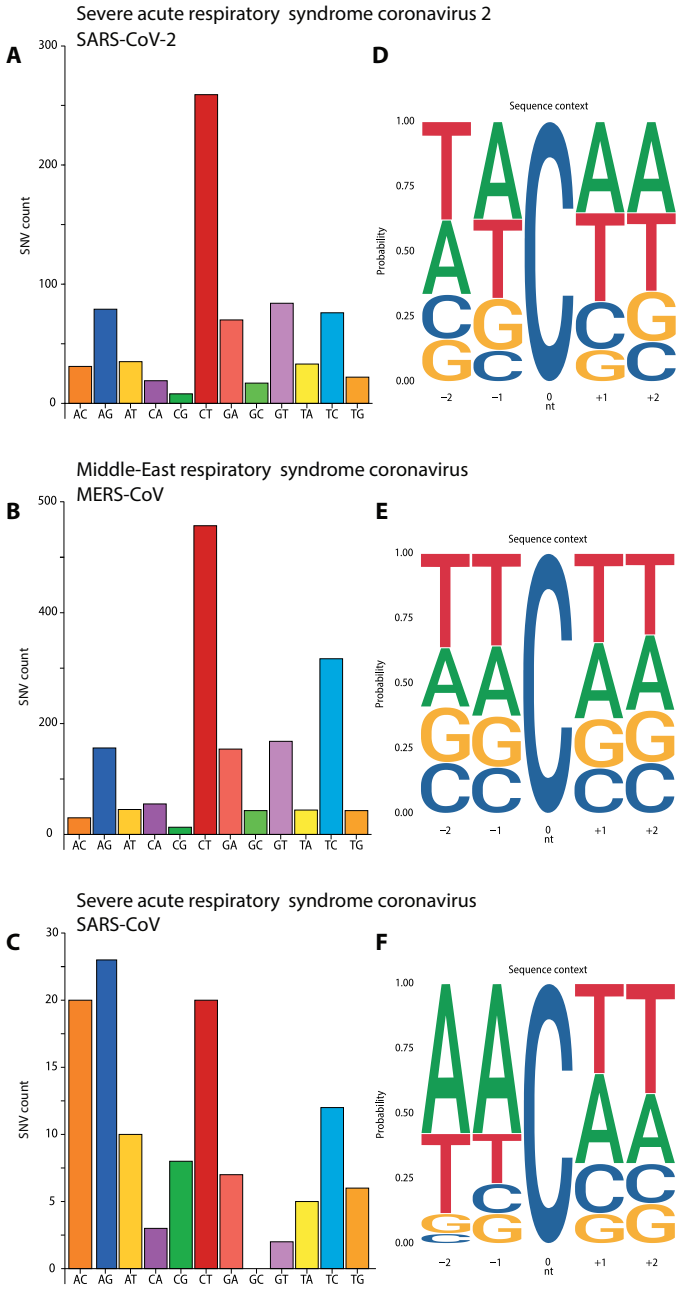

Fig. 4. Nucleotide changes across Coronaviridae strains. (**A** to **C**) Number of SNVs for each nucleotide change and (**D** to **F**) local sequence context for C-to-U edited sites in genome alignments from SARS-CoV-2 (A and D), human-hosted MERS-CoV (B and E), and human-hosted SARS-CoV (C and F).

SNVs could be due to an imbalance toward positive-sense transcripts, as these are continuously generated from the negative-sense ones (and double-stranded hybrid RNAs are lost). However, the editing frequencies of G>A SNVs should be much higher, as G>A SNVs are generated upstream to the C>T ones. A more fitting explanation is that editing of the negative-sense transcripts results in loss of the edited transcript (Fig. 5D), possibly because editing triggers nonsense-mediated decay (*55*), thus lowering the chances of the edited site to be transmitted.

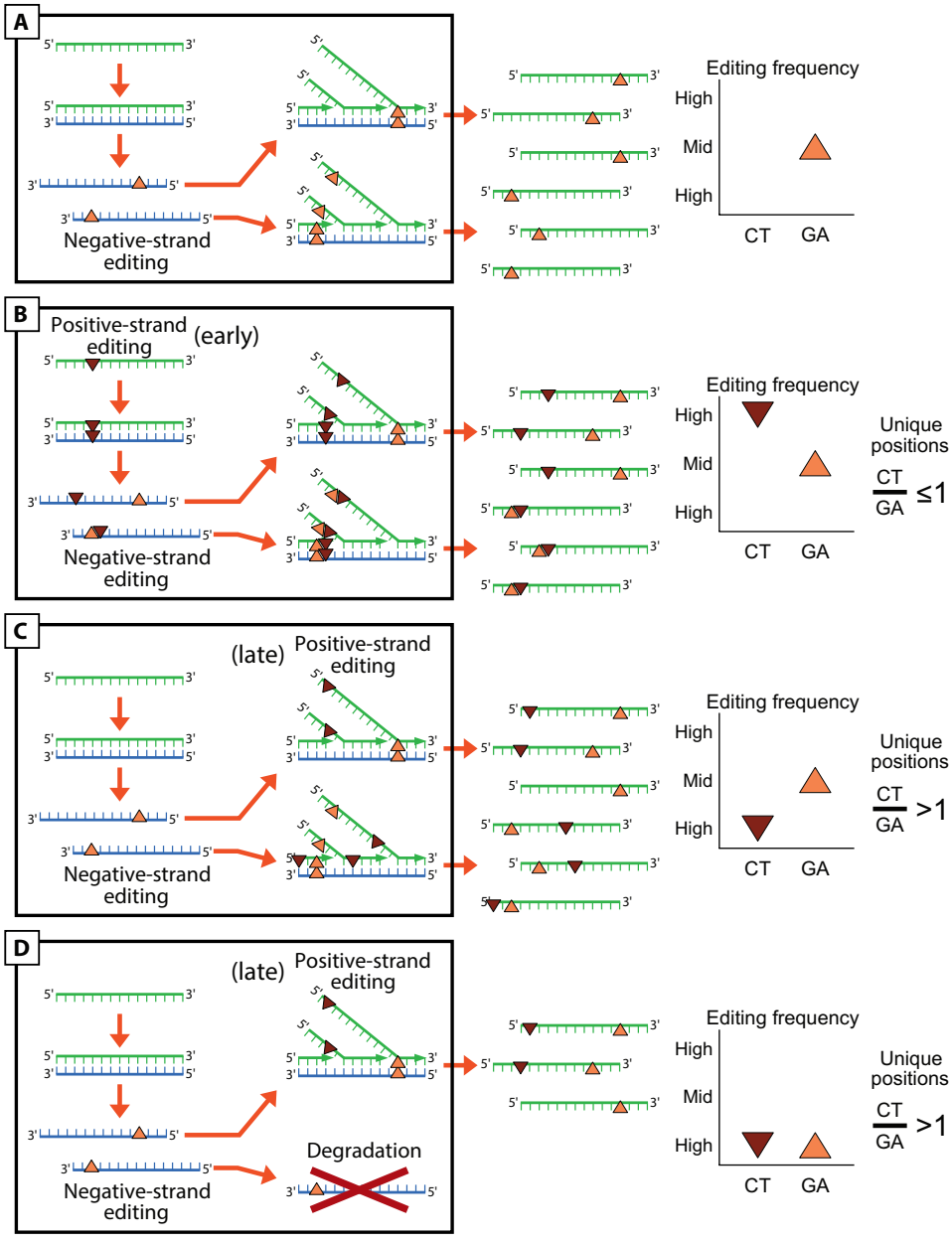

Fig. 5. Model of APOBEC RNA editing on SARS-CoV-2 transcriptome. The four panels model the editing frequencies and the C>U/G/A ratios expected from four different scenarios: (**A**) C-to-U editing on the negative-sense transcripts, (**B**) "early" editing on the viral genomes before viral replication, (**C**) "late" editing after viral replication, and (**D**) "late" editing after viral replication with loss of negative-sense transcripts. Red dots indicate editing on the positive-sense transcript; orange dots indicate editing on the positive-sense transcript. Green and blue segments indicate positive- and negative-sense viral transcripts, respectively.

Because most of the APOBECs are unable to target RNA, the only well-characterized cytidine-targeting deaminases are APOBEC1, mainly expressed in the gastrointestinal tract, and APOBEC3A (*56*), whose physiological role is not clear. As with A-to-I editing, it will be important to assess the true extent of APOBEC RNA editing in infected cells.

The functional meaning of RNA editing in SARS-CoV-2 is yet to be understood: In other contexts, editing of the viral genome determines its demise or fuels its evolution. For DNA viruses, the selection is indirect, as genomes evolve to reduce potentially harmful editable sites [e.g., (*18*)], but for RNA viruses, this pressure is even stronger, as RNA editing directly affects the genetic information and efficiently edited sites disappear.

A comparison of the SNV datasets from the transcriptomic and genomic analyses reveals a different weight of A-to-I and C-to-U changes (Figs. 2B and 4A), with an underrepresentation of A-to-I in the viral genomes. As our analysis underestimates the amount of editing due to the strict parameters used, the underrepresentation of A-to-I changes could be explained by the possibility that A-to-I editing is more effective in restricting viral propagation, thus reducing the number of viral progeny showing evidence of these changes. In contrast, the remnants of less effective C-to-U editing are retained in viral progeny and get fixed during viral adaptation.

An analysis of mutation outcomes is difficult due to the low numbers of events collected so far, but there are some possibly suggestive trends (data S2). C-to-U changes leading to stop codons are overrepresented in the transcriptomic data but—as expected—disappear in the genomic dataset. This might point—again—to an antiviral role for these editing enzymes. There is also an underrepresentation of C>T missense mutations, but its meaning is difficult to interpret.

Last, this analysis is a first step in understanding the involvement of RNA editing in viral replication, and it could lead to clinically relevant outcomes: (i) If these enzymes are relevant in the host response to coronavirus infection, a deletion polymorphism quite common in the Chinese population, encompassing the end of *APOBEC3A* and most of *APOBEC3B* (*57*, *58*), could play a role in the spread of the infection. (ii) Because RNA editing and selection act orthogonally in the evolution of the viruses, comparing genomic sites that are edited with those that are mutated could lead to the selection of viral regions potentially exploitable for therapeutic uses.

MATERIALS AND METHODS

Sequencing data

RNA sequencing data available from projects PRJNA601736, PRJNA603194, and PRJNA605907 were downloaded from the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/sra/) using the FASTQ-dump utilities from the SRA-toolkit with the following command line:

```
prefetch -v SRR* && fastq-dump --outdir /path_dir/| --split-files /path_dir/SRR*.sra
```

Because most of the reads of samples from PRJNA605907 were missing their mate, forward-reads and reverse-reads from these samples have been merged in a single FASTQ, which is treated as a single-end experiment. Details of the sequencing runs are summarized in data S1.

Data preprocessing

SRR11059940, SRR11059941, SRR11059942, and SRR11059945 showed a reduced quality of the sequencing in the terminal part of the reads. We used TRIMMOMATIC (*59*) to trim the reads of those samples to 100 base pairs (bp), with the following command line:

```
rimmomatic SE SRR*.fastq SRR*.trimmed.fastq CROP:100
```

We aligned the FASTQ files using Burrows-Wheeler Aligner (*60*) using the official sequence of SARS-CoV-2 (NC_045512. 2) as reference genome. After the alignments, BAM files were sorted using SAMtools (*61*).

The command line used for paired-end samples is as follows:

```
bwa mem NC_045512.2.fa SRR*_1.fastq SRR*_2.fastq | samtools sort -O BAM -o SRR*_.bam
```

The command line used for single-end samples is as follows:

```
bwa mem NC_045512.2.fa SRR*.fastq | samtools sort -O BAM -o SRR*_.bam
```

The aligned bams have been analyzed with QUALIMAP (*62*). Because of a high error rate reported by QUALIMAP, samples SRR11059943 and SRR10971381 have been removed from the analysis.

SNV calling

A diagram of the entire pipeline is shown in fig. S1A. We used REDItools 2 (*35*, *37*) and JACUSA (*38*) to call the SNVs using the following command line:

```
python2.7 reditools.py -f SRR*.bam -o SRR10903401_stat_table_allPos.txt -S -s 0 -os 4 -m /homol_site/SRR*_homopol.txt -c SRR*_homopol.txt -r /Reference/NC_045512.2.fa -a SRR*_stat_table_allPos.txt -q 25 -bq 35 -mbp 15 -Mbp 15
jacusa call-1 -p 20 -r SRR*.vcf -a B,I,Y -s -f V -q 35 -m 25 SRR*.srt.bam
```

With regard to REDItools 2, we removed all SNVs within 15 nucleotides from the beginning or the end of the reads to avoid artifacts due to misalignments.

To avoid potential artifacts due to strand bias, we used the AS_StrandOddsRatio parameter, calculated following GATK guidelines (https://gatk.broadinstitute.org/hc/en-us/articles/360040507111-AS-StrandOddsRatio), and any mutation with an AS_StrandOddsRatio > 4 has been removed from the dataset.

Bcftools (*61*) has been used to calculate total allelic depths on the forward and reverse strand (ADF and ADR) for AS_StrandOddsRatio calculation, with the following command line:

```
mpileup -a FORMAT/AD,FORMAT/ADF,FORMAT/ADR,FORMAT/DP,FORMAT/SP -O v -A -C -I -d 1000000 -q 25 -Q 35 -f NC_045512.2.fa -o SRR*.vcf SRR*.srt.bam
```

Mutations common to the datasets generated by REDItools 2 and JACUSA were considered ($n = 910$; fig. S2 and data S3). The threshold we used to filter the SNVs is based on minimum coverage (20 reads), number of supporting reads (at least four mutated reads), allelic fraction (0.5%), quality of the mapped reads (>25), and base quality (>35). In the dataset, there were only six SNVs with allelic fractions in the range of 30 to 85% (C>T, 1; T>C, 3; G>T, 2). Because there were no SNVs with higher allelic fractions, we presume that all samples originated from the same viral strain.

Recurring SNVs have been defined as the SNVs present in at least two samples. To overcome the problem of samples with lower sequencing depth, we used the positions of the SNVs common to both REDItools 2 and JACUSA to call again the SNVs irrespectively of the number of supporting reads.

Data manipulation

R packages (Biostrings, rsamtools, ggseqlogo ggplot2, and splitstackshape) and custom Perl scripts were used to handle the data.

Sequence context analysis

Logo alignments were calculated using ggseqlogo, using either the pooled dataset or the dataset of recurring SNVs. Logo alignments of the human edited sites were performed using ADAR sites from REDIportal (*48*) that were shared by at least four samples. SARS-CoV-2, SARS,

and MERS genomic data were prepared for the Logi alignment using the GenomicRanges R package (*63*).

SNV calling in genomic data from SARS-CoV-2, SARS, and MERS

The viral genomic sequences of MERS (taxid:1335626) and SARS (taxid:694009) were selected on NCBI Virus (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/) using the following query: Host : Homo Sapiens (human), taxid:9606; -Nucleotide Sequence Type: Complete. They were aligned using the "Align" utility. Consensus sequences of SARS and MERS genomes were built using the "cons" tool from the EMBOSS suite (http://bioinfo.nhri.org.tw/gui/) with default settings. SARS-CoV-2 genomic sequences were downloaded from GISAID (https://www.gisaid.org/) and aligned with MUSCLE (*64*).

SNVs have been called with a custom R script, by comparing viral genome sequences to the respective consensus sequence or, for SARS-CoV-2, to the NC_045512.2 reference sequence. SNVs, viral consensus sequences, and Coronaviridae genome sequence identifiers are provided in data S3 to S5.

SNV annotation

SNVs (from both genomic and somatic SNV sets) occurring on coding sequences have been annotated with custom R scripts to determine the outcome of the nucleotide change (nonsense/missense/synonymous mutation). A summary is reported in data S2.

Statistical analysis

fisher.test() function from the R base package has been used for all the statistical tests. To test the significance of C-to-U bias on the positive strand, we compared C>T/G>A SNV counts to the count of C/G bases on the reference genome. For *P* values of "RNA vs Reference," "DNA vs Reference," and "genome vs RNA," 2 × 2 contingency tables have been generated as shown in data S2.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/25/eabb5813/DC1

REFERENCES AND NOTES

1. F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
2. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Y. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, Z. Feng, Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207 (2020).
3. M. A. O'Connell, N. M. Mannion, L. P. Keegan, The epitranscriptome and innate immunity. *PLOS Genet.* **11**, e1005687 (2015).
4. E. Eisenberg, E. Y. Levanon, A-to-I RNA editing—Immune protector and transcriptome diversifier. *Nat. Rev. Genet.* **19**, 473–490 (2018).
5. R. S. Harris, J. P. Dudley, APOBECs and virus restriction. *Virology* **479–480**, 131–145 (2015).
6. J. D. Salter, H. C. Smith, Modeling the embrace of a mutator: APOBEC selection of nucleic acid ligands. *Trends Biochem. Sci.* **43**, 606–622 (2018).
7. D. R. Taylor, M. Puig, M. E. R. Darnell, K. Mihalik, S. M. Feinstone, New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. *J. Virol.* **79**, 6291–6298 (2005).
8. R. C. Zahn, I. Schelp, O. Utermöhlen, D. von Laer, A-to-G hypermutation in the genome of lymphocytic choriomeningitis virus. *J. Virol.* **81**, 457–464 (2007).
9. A. Phuphuakrat, R. Kraiwong, C. Boonarkart, D. Lauhakirti, T.-H. Lee, P. Auewarakul, Double-stranded RNA adenosine deaminases enhance expression of human immunodeficiency virus type 1 proteins. *J. Virol.* **82**, 10864–10872 (2008).
10. J. A. Carpenter, L. P. Keegan, L. Wilfert, M. A. O'Connell, F. M. Jiggins, Evidence for ADAR-induced hypermutation of the Drosophila sigma virus (Rhabdoviridae). *BMC Genet.* **10**, 75 (2009).
11. M. Doria, F. Neri, A. Gallo, M. G. Farace, A. Michienzi, Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. *Nucleic Acids Res.* **37**, 5848–5858 (2009).
12. C. X. George, Z. Li, K. M. Okonski, A. M. Toth, Y. Wang, C. E. Samuel, Tipping the balance: Antagonism of PKR kinase and ADAR1 deaminase functions by virus gene products. *J. Interferon Cytokine Res.* **29**, 477–487 (2009).
13. G. Clerzius, J.-F. Gélinas, A. Gatignol, Multiple levels of PKR inhibition during HIV-1 replication. *Rev. Med. Virol.* **21**, 42–53 (2011).
14. M. Doria, S. Tomaselli, F. Neri, S. A. Ciafrè, M. G. Farace, A. Michienzi, A. Gallo, ADAR2 editing enzyme is a novel human immunodeficiency virus-1 proviral factor. *J. Gen. Virol.* **92**, 1228–1232 (2011).
15. C. K. Pfaller, Z. Li, C. X. George, C. E. Samuel, Protein kinase PKR and RNA adenosine deaminase ADAR1: New roles for old players as modulators of the interferon response. *Curr. Opin. Immunol.* **23**, 573–582 (2011).
16. C. E. Samuel, ADARs: Viruses and innate immunity. *Curr. Top. Microbiol. Immunol.* **353**, 163–195 (2012).
17. S. Tomaselli, F. Galeano, F. Locatelli, A. Gallo, ADARs and the balance game between virus infection and innate immune cell response. *Curr. Issues Mol. Biol.* **17**, 37–51 (2015).
18. U. Rosani, C.-M. Bai, L. Maso, M. Shapiro, M. Abbadi, S. Domeneghetti, C.-M. Wang, L. Cendron, T. MacCarthy, P. Venier, A-to-I editing of *Malacoherpesviridae* RNAs supports the antiviral role of ADAR1 in mollusks. *BMC Evol. Biol.* **19**, 149 (2019).
19. J. P. Vartanian, A. Meyerhans, B. Asjö, S. Wain-Hobson, Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.* **65**, 1779–1788 (1991).
20. R. S. Harris, K. N. Bishop, A. M. Sheehy, H. M. Craig, S. K. Petersen-Mahrt, I. N. Watt, M. S. Neuberger, M. H. Malim, DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803–809 (2003).
21. R. Mahieux, R. Suspène, F. Delebecque, M. Henry, O. Schwartz, S. Wain-Hobson, J.-P. Vartanian, Extensive editing of a small fraction of human T-cell leukemia virus type 1 genomes by four APOBEC3 cytidine deaminases. *J. Gen. Virol.* **86**, 2489–2494 (2005).
22. C. Noguchi, H. Ishino, M. Tsuge, Y. Fujimoto, M. Imamura, S. Takahashi, K. Chayama, G to A hypermutation of hepatitis B virus. *Hepatology* **41**, 626–633 (2005).
23. J.-P. Vartanian, D. Guétard, M. Henry, S. Wain-Hobson, Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* **320**, 230–233 (2008).
24. R. Suspène, M.-M. Aynaud, S. Koch, D. Pasdeloup, M. Labetoulle, B. Gaertner, J.-P. Vartanian, A. Meyerhans, S. Wain-Hobson, Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and *in vivo*. *J. Virol.* **85**, 7594–7602 (2011).
25. J. M. Cuevas, R. Geller, R. Garijo, J. López-Aldeguer, R. Sanjuán, Extremely high mutation rate of HIV-1 in vivo. *PLOS Biol.* **13**, e1002251 (2015).
26. A. Peretti, E. M. Geoghegan, D. V. Pastrana, S. Smola, P. Feld, M. Sauter, S. Lohse, M. Ramesh, E. S. Lim, D. Wang, C. Borgogna, P. C. FitzGerald, V. Bliskovsky, G. J. Starrett, E. K. Law, R. S. Harris, J. K. Killian, J. Zhu, M. Pineda, P. S. Meltzer, R. Boldorini, M. Gariglio, C. B. Buck, Characterization of BK polyomaviruses from kidney transplant recipients suggests a role for APOBEC3 in driving in-host virus evolution. *Cell Host Microbe* **23**, 628–635.e7 (2018).
27. E. N. C. Newman, R. K. Holmes, H. M. Craig, K. C. Klein, J. R. Lingappa, M. H. Malim, A. M. Sheehy, Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr. Biol.* **15**, 166–170 (2005).
28. D. Pollpeter, M. Parsons, A. E. Sobala, S. Coxhead, R. D. Lang, A. M. Bruns, S. Papaioannou, J. M. McDonnell, L. Apolonia, J. A. Chowdhury, C. M. Horvath, M. H. Malim, Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. *Nat. Microbiol.* **3**, 220–233 (2018).
29. A. Milewska, E. Kindler, P. Vkovski, S. Zeglen, M. Ochman, V. Thiel, Z. Rajfur, K. Pyrc, APOBEC3-mediated restriction of RNA virus replication. *Sci. Rep.* **8**, 5960 (2018).
30. J. S. Albin, G. Haché, J. F. Hultquist, W. L. Brown, R. S. Harris, Long-term restriction by APOBEC3F selects human immunodeficiency virus type 1 variants with restored Vif function. *J. Virol.* **84**, 10209–10219 (2010).
31. H. A. Sadler, M. D. Stenglein, R. S. Harris, L. M. Mansky, APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J. Virol.* **84**, 7396–7404 (2010).
32. L. Perelygina, M.-H. Chen, S. Suppiah, A. Adebayo, E. Abernathy, M. Dorsey, L. Bercovitch, K. Paris, K. P. White, A. Krol, J. Dhossche, I. Y. Torshin, N. Saini, L. J. Klimczak, D. A. Gordenin, A. Zharkikh, S. Plotkin, K. E. Sullivan, J. Icenogle, Infectious vaccine-derived rubella viruses emerge, persist, and evolve in cutaneous granulomas of children with primary immunodeficiencies. *PLOS Pathog.* **15**, e1008080 (2019).
33. L. Chen, W. Liu, Q. Zhang, K. Xu, G. Ye, W. Wu, Z. Sun, F. Liu, K. Wu, B. Zhong, Y. Mei, W. Zhang, Y. Chen, Y. Li, M. Shi, K. Lan, Y. Liu, RNA based mNGS approach identifies

a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. *Emerg. Microbes Infect.* **9**, 313–319 (2020).
34. Z. Shen, Y. Xiao, L. Kang, W. Ma, L. Shi, L. Zhang, Z. Zhou, J. Yang, J. Zhong, D. Yang, L. Guo, G. Zhang, H. Li, Y. Xu, M. Chen, Z. Gao, J. Wang, L. Ren, M. Li, Genomic diversity of SARS-CoV-2 in coronavirus disease 2019 patients. *Clin. Infect. Dis.* , ciaa203 (2020).
35. E. Picardi, G. Pesole, REDItools: High-throughput RNA editing detection made easy. *Bioinformatics* **29**, 1813–1814 (2013).
36. C. Lo Giudice, M. A. Tangaro, G. Pesole, E. Picardi, Investigating RNA editing in deep transcriptome datasets with REDItools and REDIportal. *Nat. Protoc.* **15**, 1098–1131 (2020).
37. T. Flati, S. Gioiosa, N. Spallanzani, I. Tagliaferri, M. Angela Diroma, G. Pesole, G. Chillemi, E. Picardi, T. Castrignanò, HPC-REDItools: A novel HPC-aware tool for improved large scale RNA-editing analysis. bioRxiv 2020.04.30.069732 [**Preprint**]. 30 April 2020. https://doi.org/10.1101/2020.04.30.069732.
38. M. Piechotta, E. Wyler, U. Ohler, M. Landthaler, C. Dieterich, JACUSA: Site-specific identification of RNA editing events from replicate sequencing data. *BMC Bioinformatics* **18**, 7 (2017).
39. L. D. Eckerle, M. M. Becker, R. A. Halpin, K. Li, E. Venter, X. Lu, S. Scherbakova, R. L. Graham, R. S. Baric, T. B. Stockwell, D. J. Spiro, M. R. Denison, Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLOS Pathog.* **6**, e1000896 (2010).
40. M. R. Denison, R. L. Graham, E. F. Donaldson, L. D. Eckerle, R. S. Baric, Coronaviruses. *RNA Biol.* **8**, 270–279 (2011).
41. E. C. Smith, H. Blanc, M. C. Surdel, M. Vignuzzi, M. R. Denison, Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: Evidence for proofreading and potential therapeutics. *PLOS Pathog.* **9**, e1003565 (2013).
42. L. Bazak, A. Haviv, M. Barak, J. Jacob-Hirsch, P. Deng, R. Zhang, F. J. Isaacs, G. Rechavi, J. B. Li, E. Eisenberg, E. Y. Levanon, A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **24**, 365–376 (2014).
43. D. Bar-Yaacov, G. Avital, L. Levin, A. L. Richards, N. Hachen, B. Rebolledo Jaramillo, A. Nekrutenko, R. Zarivach, D. Mishmar, RNA–DNA differences in human mitochondria restore ancestral form of 16S ribosomal RNA. *Genome Res.* **23**, 1789–1796 (2013).
44. K. A. Lehmann, B. L. Bass, Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875–12884 (2000).
45. J. B. Li, E. Y. Levanon, J.-K. Yoon, J. Aach, B. Xie, E. Leproust, K. Zhang, Y. Gao, G. M. Church, Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
46. J. H. Bahn, J.-H. Lee, G. Li, C. Greer, G. Peng, X. Xiao, Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–150 (2012).
47. E. Picardi, C. Manzari, F. Mastropasqua, I. Aiello, A. M. D'Erchia, G. Pesole, Profiling RNA editing in human tissues: Towards the inosinome Atlas. *Sci. Rep.* **5**, 14941 (2015).
48. E. Picardi, A. M. D'Erchia, C. Lo Giudice, G. Pesole, REDIportal: A comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **45**, D750–D757 (2017).
49. B. R. Rosenberg, C. E. Hamilton, M. M. Mwangi, S. Dewell, F. N. Papavasiliou, Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. *Nat. Struct. Mol. Biol.* **18**, 230–236 (2011).
50. T. Lerner, F. N. Papavasiliou, R. Pecori, RNA editors, cofactors, and mRNA targets: An overview of the C-to-U RNA editing machinery and its implication in human disease. *Genes* **10**, 13 (2018).
51. S. H. Roth, E. Y. Levanon, E. Eisenberg, Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity. *Nat. Methods* **16**, 1131–1138 (2019).
52. S. Osenberg, D. Dominissini, G. Rechavi, E. Eisenberg, Widespread cleavage of A-to-I hyperedited substrates. *RNA* **15**, 1632–1639 (2009).
53. N. L. Ko, E. Birlouez, S. Wain-Hobson, R. Mahieux, J.-P. Vartanian, Hyperediting of human T-cell leukemia virus type 2 and simian T-cell leukemia virus type 3 by the dsRNA adenosine deaminase ADAR-1. *J. Gen. Virol.* **93**, 2646–2651 (2012).
54. H. T. Porath, S. Carmi, E. Y. Levanon, A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.* **5**, 4726 (2014).
55. A. Chester, A. Somasekaram, M. Tzimina, A. Jarmuz, J. Gisbourne, R. O'Keefe, J. Scott, N. Navaratnam, The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J.* **22**, 3971–3982 (2003).
56. S. Sharma, S. K. Patnaik, Z. Kemer, B. E. Baysal, Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. *RNA Biol.* **14**, 603–610 (2017).
57. J. M. Kidd, T. L. Newman, E. Tuzun, R. Kaul, E. E. Eichler, Population stratification of a common *APOBEC* gene deletion polymorphism. *PLOS Genet.* **3**, e63 (2007).
58. J. Long, R. J. Delahanty, G. Li, Y.-T. Gao, W. Lu, Q. Cai, Y.-B. Xiang, C. Li, B.-T. Ji, Y. Zheng, S. Ali, X.-O. Shu, W. Zheng, A common deletion in the *APOBEC3* genes and breast cancer risk. *J. Natl. Cancer Inst.* **105**, 573–579 (2013).
59. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
60. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
61. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
62. K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
63. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
64. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

Acknowledgments: In memory of Li Wenliang, Carlo Urbani, and all the doctors and health workers who endangered their lives in the fight against epidemics. We acknowledge and thank all authors that are sharing their data. We gratefully acknowledge the authors, originating and submitting laboratories, of the sequences from GISAID's EpiFlu Database on which this research is based. The list is detailed in data S6. We thank Ernesto Picardi for the support with REDItools2. **Funding:** The research was supported by grants from Ministero della Salute (PE-2013-02357669) and from AIRC (IG-17701). **Author contributions:** Conceptualization: S.D.G., F.M., M.G.T., G.M., and S.G.C.; formal analysis, investigation, and software: S.D.G. and F.M.; visualization: G.M.; writing (original draft): G.M. and S.G.C.; writing (review and editing): S.D.G., F.M., M.G.T., G.M., and S.G.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Sequencing and genomic data are available through NCBI SRA and Virus repositories and GISAID. The SNVs dataset is also available through the UCSC genome browser at http://genome.ucsc.edu/s/Max/conticello2020.

Submitted 4 March 2020
Accepted 5 May 2020
First Release 18 May 2020
Published 17 June 2020
10.1126/sciadv.abb5813

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

Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2

Salvatore Di Giorgio, Filippo Martignano, Maria Gabriella Torcia, Giorgio Mattiuz and Silvestro G. Conticello

Sci Adv **6** (25), eabb5813.
DOI: 10.1126/sciadv.abb5813originally published online May 18, 2020