Evidence for RNA editing in the transcriptome of 2019 Novel Coronavirus

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

Affiliations:

[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.

*Correspondence to: giorgio.mattiuz@unifi.it, s.conticello@ispro.toscana.it

† ‡ These authors contributed equally

Abstract:

The 2019-nCoV outbreak has become a global health risk. Editing by host deaminases is an innate restriction process to counter viruses, and it is not yet known whether it operates against coronaviruses. Here we analyze RNA sequences from bronchoalveolar lavage fluids derived from two Wuhan 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 ones. A mutational analysis of genomes from different strains of human-hosted Coronaviridae reveals patterns similar to the RNA editing pattern observed in the 2019-nCoV transcriptomes. Our results suggest that both APOBECs and ADARs are involved in Coronavirus genome editing, a process that may shape the fate of both virus and patient.

Main Text:

Emerging viral infections represent a threat to global health, and the recent outbreak of Novel Coronavirus 2019 (2019-nCoV or SARS-CoV-2) from Wuhan (China) exemplifies the risks (*1*, *2*). As viruses are obligate intracellular parasites, organisms evolved innate immune processes for the sensing and the restriction of viruses. Among the processes involved is RNA and DNA editing mediated by endogenous deaminases. Two different deaminase families are present in mammalian species: the ADARs 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 (ssDNA and ssRNA) (*5*, *6*). ADARs interfere with viral infections directly -through hypermutation of viral RNA- and indirectly, through modulation of the intracellular response (*7*–*12*). On the other hand, APOBECs target the viral genome, typically DNA intermediates (*13*–*20*), either through C-to-U hypermutation or through a non-enzymatic path that interferes with reverse transcription (*21*, *22*). Some APOBEC3 proteins can interfere *in vitro* with *Coronaviridae* replication, yet it is not clear whether their enzymatic activity is involved (*23*). Eventually though, both of these restriction systems are exploited by the viruses themselves to increase their evolutionary potential (*24*–*26*).

To assess whether RNA editing could be involved in the response to 2019-nCoV infections, we started from publicly available RNA sequencing datasets from bronchoalveolar lavage fluids (BALF) obtained from patients diagnosed with Coronavirus Virus disease (COVID-2019). While transcriptomic data for all samples could be aligned to the 2019-nCoV reference genome, the quality of the sequencing varied among samples, and only two samples had coverage and error rates suitable for the identification of potentially edited sites (**Supplementary Table 1**). We therefore called the single nucleotide variants (SNVs) using REDItools 2 only on the data from the two patients reported in *Chen et al.* (*27*). We identified 487 SNVs, 41 from patient SRR10903401 and 446 in the patient SRR10903402 (**Fig. S1, Data S1**).

Fig. 1

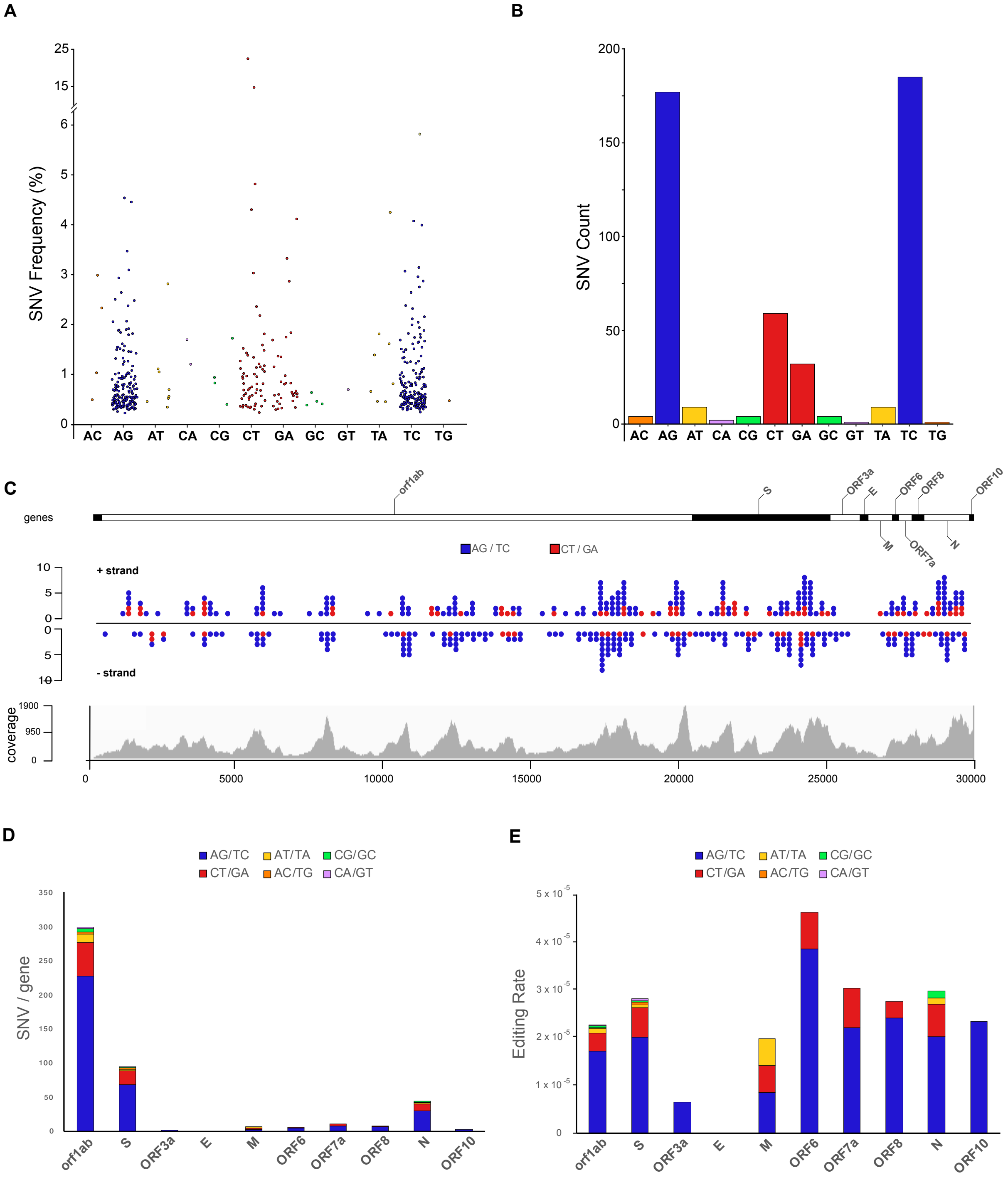


Fig. 1. Single-nucleotide variants (SNV) identified in 2019-nCoV transcriptomes. (**A**) Allelic fraction and (**B**) number of SNVs for each nucleotide change (e.g. A>C, AC). **(C)** Distribution of SNVs across the 2019-nCoV genome (middle). AG/TC (*blue*) and CT/GA (*red*). SNVs are grouped in bins of 200 nt and are plotted above (AG and CT) or below the line (TC and GA) based on the edited strand. Genetic organization of 2019-nCoV (top); coverage distribution of sample SRR10903402 (bottom). (**D**) SNV count and type per gene. (**E**) Editing rate: SNV count and type per gene after normalization by gene length and coverage.

The allelic fraction of the SNVs ranges between 1 and 5% (**Fig. 1A**) and there is a bias towards transitions (**Fig.1B**). The number of transversions is compatible with the mutation rate observed in Coronaviruses ($10^{-6/-7}$, *28*). On the other hand, the bias towards transitions resembles the pattern of SNVs observed in human transcriptomes (*29*) or in viruses (*8*, *9*, *12*), where A>G changes derive from deamination of A-to-I mediated by the ADARs. The pattern we observe encompasses all possible transitions, with A>G and T>C being the main changes, evenly represented. This is compatible with the presence in the cell, during replication, of both positive and negative strands of the viral genome. It is thus likely that also in the case of 2019-nCoV these A>G/T>C changes are due to the action of the ADARs.

C>T and G>A are the second main group of changes and could derive from APOBEC-mediated C-to-U deamination. Contrary to A-to-I, C-to-U editing is a relatively rare phenomenon in the human transcriptome (*29*) and, with regards to viruses, it has been associated only to positive-sense ssRNA Rubella virus (*26*), where C>T changes represent the predominant SNV type. In support of a role for the APOBECs in 2019-nCoV RNA editing is the observation that only A-to-I editing is present in non-vertebrate RNA viruses, where there are no known RNA-targeting

APOBECs (*9*, *12*). As with A-to-I changes, C-to-U changes seem to target equally both RNA strands.

SNVs are spread throughout the viral genome, with no apparent difference in distribution between A-to-I and C-to-U changes (**Fig. 1C**). Since ADARs and APOBECs target selectively dsRNA and ssRNA, such distribution could derive 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. Some areas seem to bear less SNVs, but this is due to lower coverage.

Similarly, SNVs accumulate on most viral genes proportionally to gene length and sequencing depth (**Fig. 1D, E**). Only exception is the ORF6 gene, which modulates interferon signaling (*30*), where the editing rate is higher compared to that of the other genes.

Since APOBEC deaminases preferentially target cytosines within specific sequence contexts, we analyzed the trinucleotide context of the SNVs. After normalization for the genomic context, no apparent bias is visible for A-to-I changes (**Fig. 2A, B**). On the other hand, C-to-U changes preferentially occur 3' to thymines and adenines, as logo alignment clearly points towards a pattern compatible with APOBEC1 deamination ([AU]C[AU] (**Fig. 2B**) (*31*, *32*). Moreover, the most frequently edited trinucleotide context is TCG (**Fig. 2A**), which is compatible with APOBEC3A pattern (UCN) (*33*).

Fig. 2

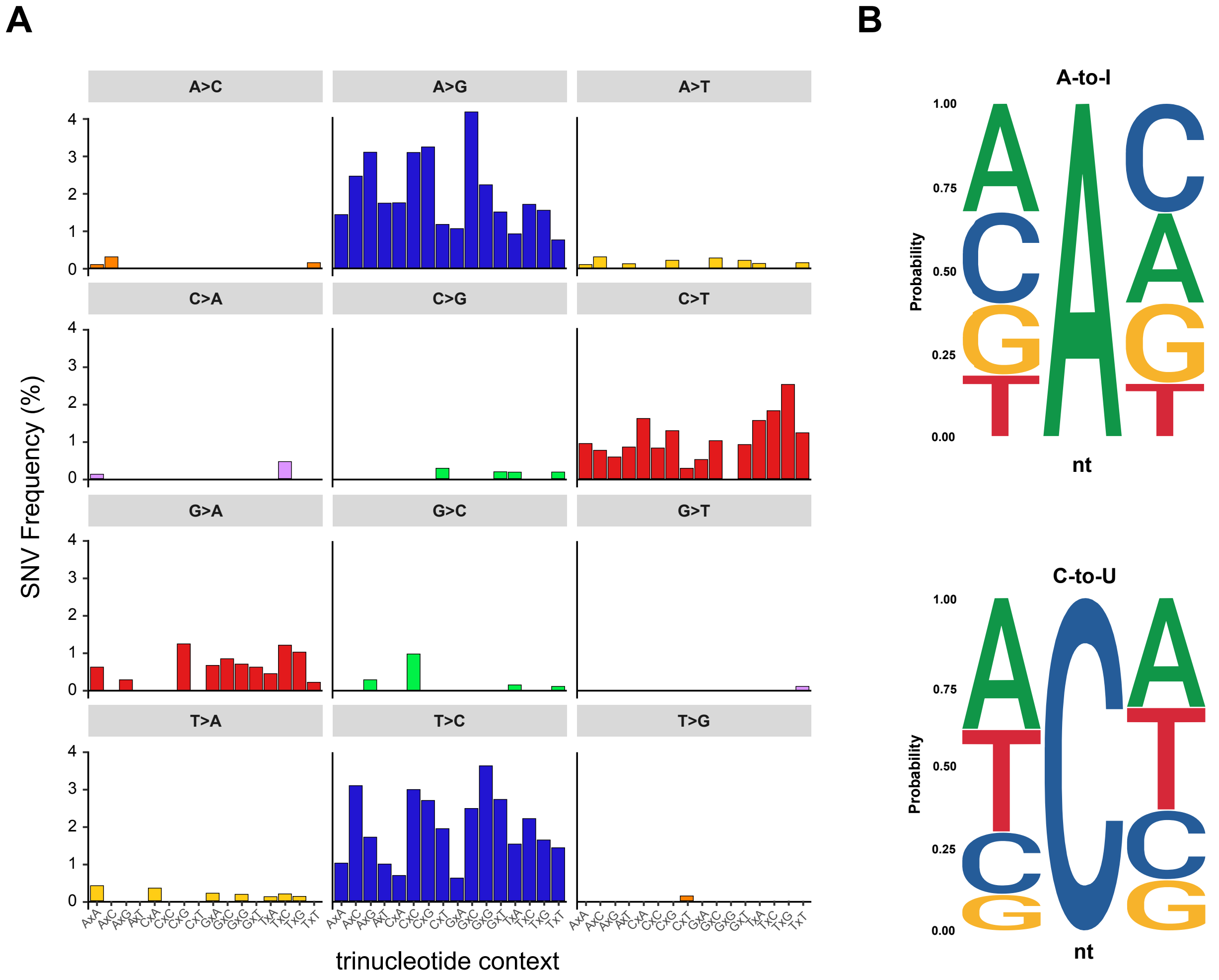


Fig. 2. Sequence contexts for 2019-nCoV RNA edited sites. A, 2019-nCoV SNV frequency in each trinucleotide context (5' to 3', with x indicating the edited position) after normalization by genomic content. Frequencies are normalized on the frequency of each trinucleotide context across the viral genome (see methods). **B,** Local sequence context for A-to-I and C-to-U edited sites in the viral transcriptome.

We then aligned available Coronavirus genomes from Novel Coronavirus 2019 (2019-nCoV), Middle East respiratory syndrome-related coronavirus (MERS-CoV), and the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) to test whether RNA editing could be responsible for some of the mutations acquired through evolution. Indeed, the genomic alignments reveal that a substantial fraction of the mutations in all strains could derive from C-to-U and A-to-I deamination (**Fig. 3A, C, E**) and that a pattern compatible with APOBEC-mediated editing exists also in genomic C-to-U SNVs (**Fig. 3B, D, F**).

Fig. 3

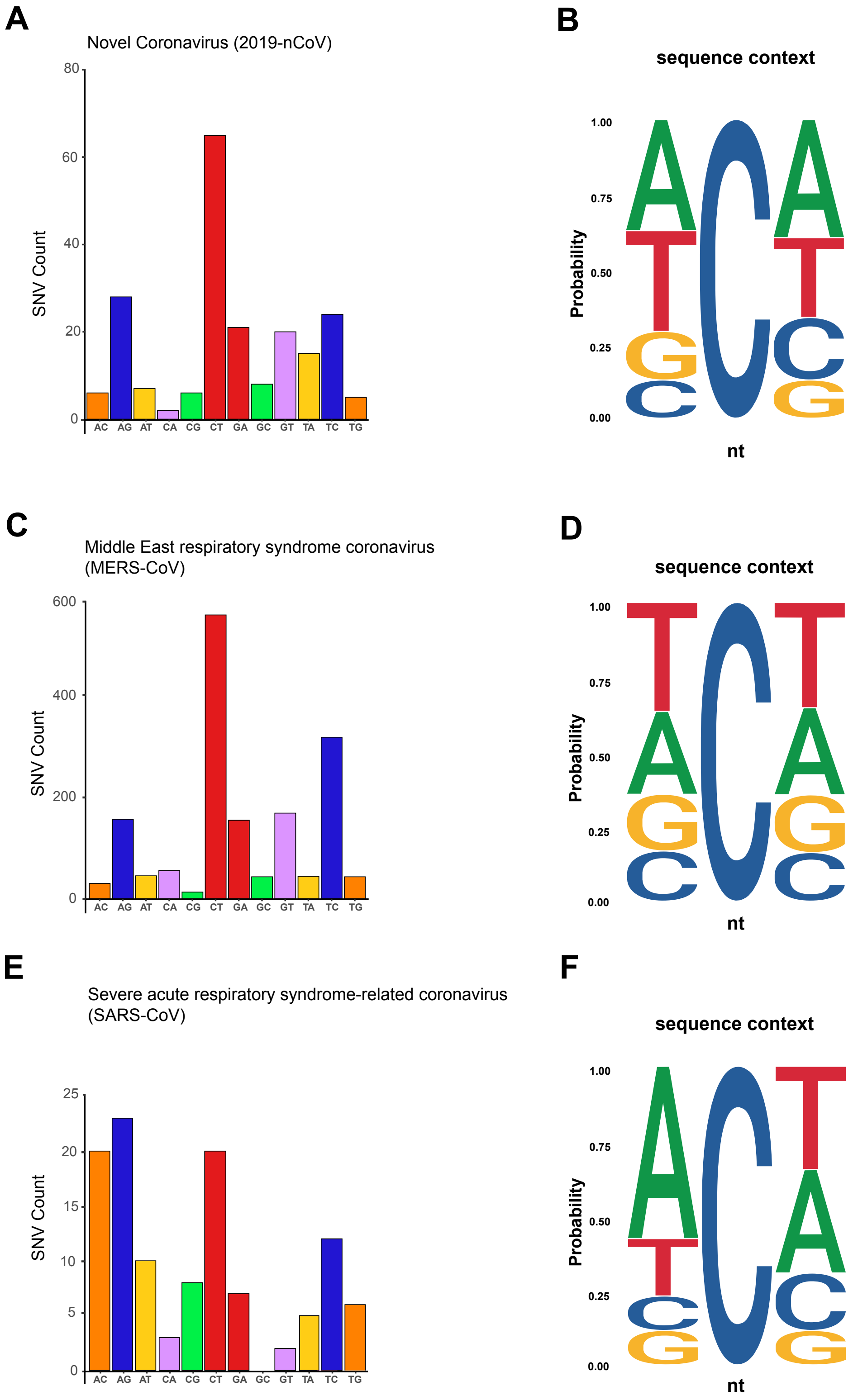

A
Novel Coronavirus (2019-nCoV)
SNV Count
0
20
40
60
80
AC
AG
AT
CA
CG
CT
GA
GC
GT
TA
TC
TG
B
sequence context
Probability
0.00
0.25
0.50
0.75
1.00
nt
C
Middle East respiratory syndrome coronavirus (MERS-CoV)
SNV Count
0
200
400
600
AC
AG
AT
CA
CG
CT
GA
GC
GT
TA
TC
TG
D
sequence context
Probability
0.00
0.25
0.50
0.75
1.00
nt
E
Severe acute respiratory syndrome-related coronavirus (SARS-CoV)
SNV Count
0
5
10
15
20
25
AC
AG
AT
CA
CG
CT
GA
GC
GT
TA
TC
TG
F
sequence context
Probability
0.00
0.25
0.50
0.75
1.00
nt

Fig. 3. Nucleotide changes across *Coronaviridae* strains. A, C, E, Raw number of SNVs for each nucleotide change across 2019-nCoV (**A**), human-hosted MERS-CoV (**C**), and human-hosted SARS-CoV (**E**) genome alignments (**Data S2-S3**). **B, D, F,** Local sequence context for C-to-U edited sites 2019-nCoV (**B**), human-hosted MERS-CoV (**D**), and human-hosted SARS-CoV (**F**) (**Fig. S2**).

Considering the source of our analysis -metagenomic sequencing- we wonder whether the editing frequencies we observe (~1%) reflect the levels of editing of the viral RNAs inside the cell. These are the same editing levels observed for most of the ADAR-edited sites in the human transcriptome (typically inside Alu sequences) (*3*, *29*, *34*). To understand whether A-to-I editing is an effective path of restriction of 2019-nCoV, it will be necessary to assess if a substantial fraction of viral transcripts is hyper-edited (*35–37*) or fail to be packaged into virions.

With regards 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 (*29*). Such enrichment points either towards an induction of the APOBECs triggered by the infection, or to specific targeting of the APOBECs onto the viral transcripts. The 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 (*13–20*) through DNA editing- and in Rubella virus, on RNA (*26*). Intriguingly, C>T changes are predominant in Rubella virus, suggesting that the positive-sense strand is targeted. This difference between APOBEC editing in Rubella virus and in 2019-nCoV might derive from differences in the replication processes or to RNA accessibility.

Since most of the APOBECs are unable to target RNA, the only well characterized cytidine-targeting deaminase are APOBEC1, mainly expressed in the gastrointestinal tract, and APOBEC3A, 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 2019-nCoV 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. *12*), but for RNA viruses this pressure is even stronger, as RNA editing directly effects the genetic information and efficiently edited sites disappear.

Finally, this analysis is a first step in understanding the involvement of RNA editing in viral replication, and it could lead to clinically relevant outcomes: (a) if these enzymes are relevant in the host response to Coronavirus infection, a polymorphism quite common in the Chinese population that inactivates APOBEC3A (*38*, *39*) could play a role in the spread of the infection. (b) Since 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 and PRJNA603194 were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/sra/) using the fastq-dump utilities from the SRA-toolkit with the command line:

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

Details of the sequencing runs are summarised in **Table S1**.

Data pre-processing

We aligned the FASTQ files using Burrows-Wheeler Aligner (BWA) (*40*) using the official sequence of 2019-nCoV (NC_045512. 2) as reference genome.

The command line for the alignment and the sorting with Samtools (*41*) is:

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

Somatic nucleotide variant (SNV) calling

We used RediTools2 (*42*, *43*) to call the SNVs in RNA mode using the command line:

```
python2.7 reditools.py -f SRR*.bam -S -s 0 -os 4 -r NC_045512.2.fa -m omopolymeric_file.txt -c omopolymeric_file.txt -q 20 -bq 25 -o SRR*_output_table.txt
```

The threshold we used to filter the SNVs is based on minimum coverage (20 reads) and on the number of supporting reads (at least 4 mutated reads). These filters allowed us to construct a high confidence set of RNA SNVs for both patients (**Fig. S1, S2**). The SNVs identified in patients SRR10903401 and SRR10903402 are listed in **Data S1**. A diagram of the entire pipeline is shown in **Fig. S3**.

Data manipulation and analysis

R packages (*44–48*) and custom Perl scripts were used to handle the data.

Normalization of SNV counts on gene length and sequencing depth

Higher gene lengths lead to higher probability of being mutated; hence, to investigate RNA editing enrichments on 2019-nCoV genes we normalized SNV raw counts on gene-length. Since SNV raw counts correlate with sequencing depth (**Fig. S4**), we normalized SNV raw counts on genes' average sequencing depth.

Consequently, editing rate per gene has been calculated as follows:

$$Editing\ rate = \frac{N}{L \times D}$$

Where *N* is the SNV raw count occurring in a specific viral gene, *L* is the gene length, and *D* is the sum of the sequencing depth of the two samples.

Normalization of trinucleotide contexts

The analysis of the trinucleotide contexts was performed normalizing the mutated trinucleotide with respect to its presence in the 2019-nCoV genome as follows:

$$SNV\ frequency\ (\%) = \frac{C}{G}$$

where *C* is the count of SNVs occurring with a specific trinucleotide context, and *G* is the number of times that trinucleotide context is present in the viral genome.

Viral genomes alignments

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 query: Host : Homo Sapiens (human), taxid:9606; -Nucleotide Sequence Type: Complete. They were aligned using the “Align” utility.

2019-nCoV genomic sequences were downloaded from GISAID (https://www.gisaid.org/) and aligned with MUSCLE (*49*). Consensus sequences of 2019-nCoV, SARS and MERS genomes were built using the “cons” tool from the EMBOSS suite (http://bioinfo.nhri.org.tw/gui/) with default settings. SNVs have been called with a custom R script, by comparing viral genome sequences to the respective consensus sequence. Viral consensus sequences, sequences identifiers, and mutation files are provided in **Data S2**, **S3** and **Table S2**.

Fig. S1

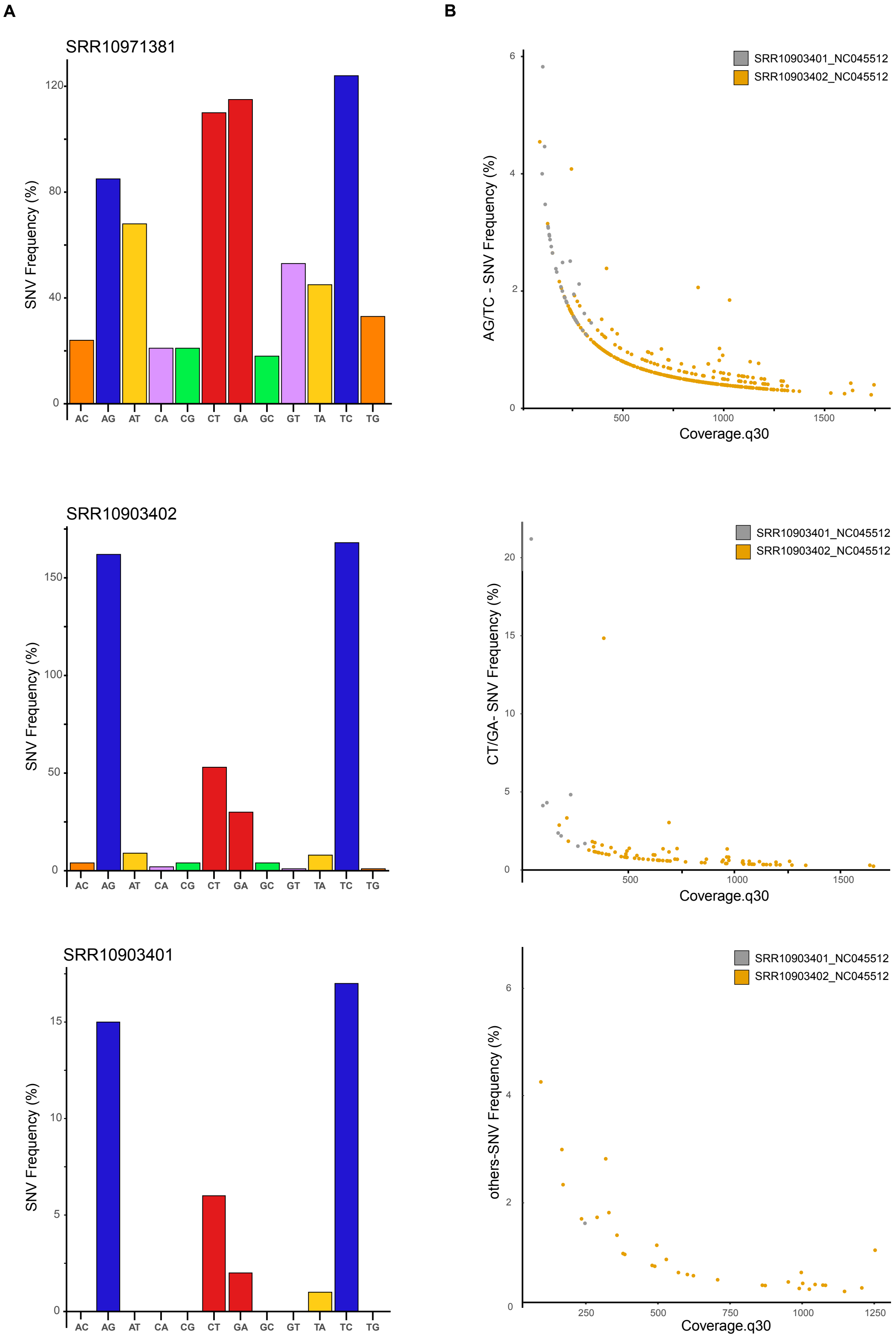

A
B
SRR10971381
SNV Frequency (%)
0
40
80
120
AC AG AT CA CG CT GA GC GT TA TC TG
SRR10903402
SNV Frequency (%)
0
50
100
150
AC AG AT CA CG CT GA GC GT TA TC TG
SRR10903401
SNV Frequency (%)
0
5
10
15
AC AG AT CA CG CT GA GC GT TA TC TG
AG/TC - SNV Frequency (%)
0
2
4
6
500
1000
1500
Coverage.q30
SRR10903401_NC045512
SRR10903402_NC045512
CT/GA- SNV Frequency (%)
0
5
10
15
20
500
1000
1500
Coverage.q30
SRR10903401_NC045512
SRR10903402_NC045512
others-SNV Frequency (%)
0
2
4
6
250
500
750
1000
1250
Coverage.q30
SRR10903401_NC045512
SRR10903402_NC045512

Fig. S1.

RNA-editing on 2019-nCoV transcriptome patients. A, Single-nucleotide variants (SNVs) identified in the RNA-sequencing for each sample (SRR10903401, SRR10903402 and SRR10971381). The bar charts (*x-axes*) represent all possible nucleotide changes. The *y-axes* represent the count of SNVs found in each 2019-nCoV transcriptome. The *y-axes* represent the count of SNVs. **B,** ADARs and APOBECs mutation frequency associated with the coverage obtained from the best RNA-sequencing derived from two Wuhan (China) patients. Each panel shows respectively: AG/TC, CT/GA and others (see Material and Methods). The *blue* and *yellow* dots represent the SNVs of each patient. The *x-axes* represent the coverage of the mapped reads ; the *y-axes* the mutation frequency.

Fig. S2

Novel Coronavirus (2019-nCoV)

	A	C	G	T
A		6	28	7
C	2		6	65
G	21	8		20
T	15	24	5	

Middle East respiratory syndrome coronavirus (MERS-CoV)

	A	C	G	T
A		30	156	45
C	55		13	557
G	154	43		168
T	44	317	43	

Severe acute respiratory syndrome-related coronavirus (SARS-CoV)

	A	C	G	T
A		20	23	10
C	3		8	20
G	7	0		2
T	5	12	6	

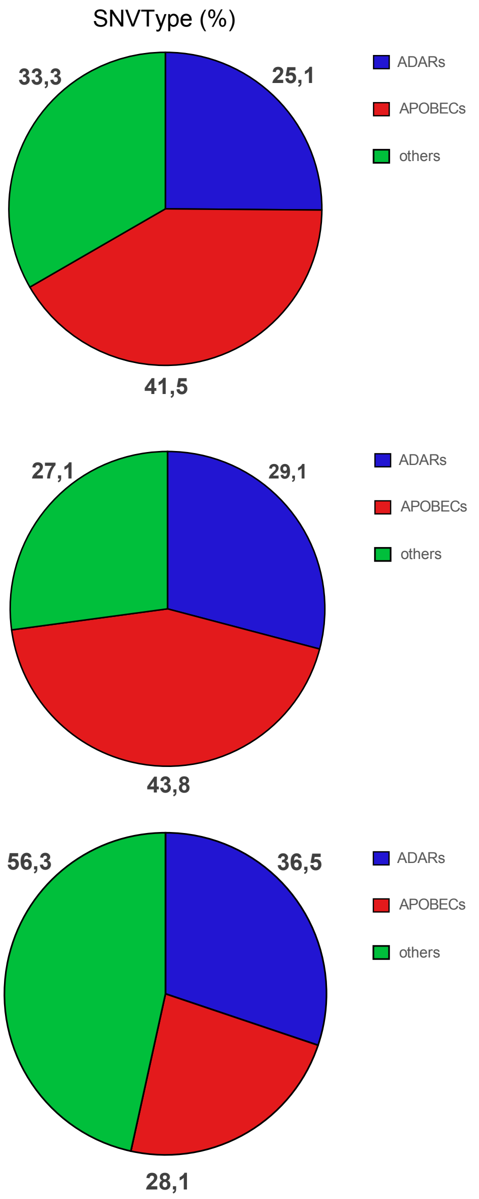


Fig. S2.

SNVs in *Coronaviridae* strains. All possible nucleotide changes that occur in 2019-nCoV, MERS and SARS genomes were reported in the corresponding table. The pies represent the percentage (%) of ADAR and APOBEC SNVs with respect to all the other SNVs.

Fig. S3

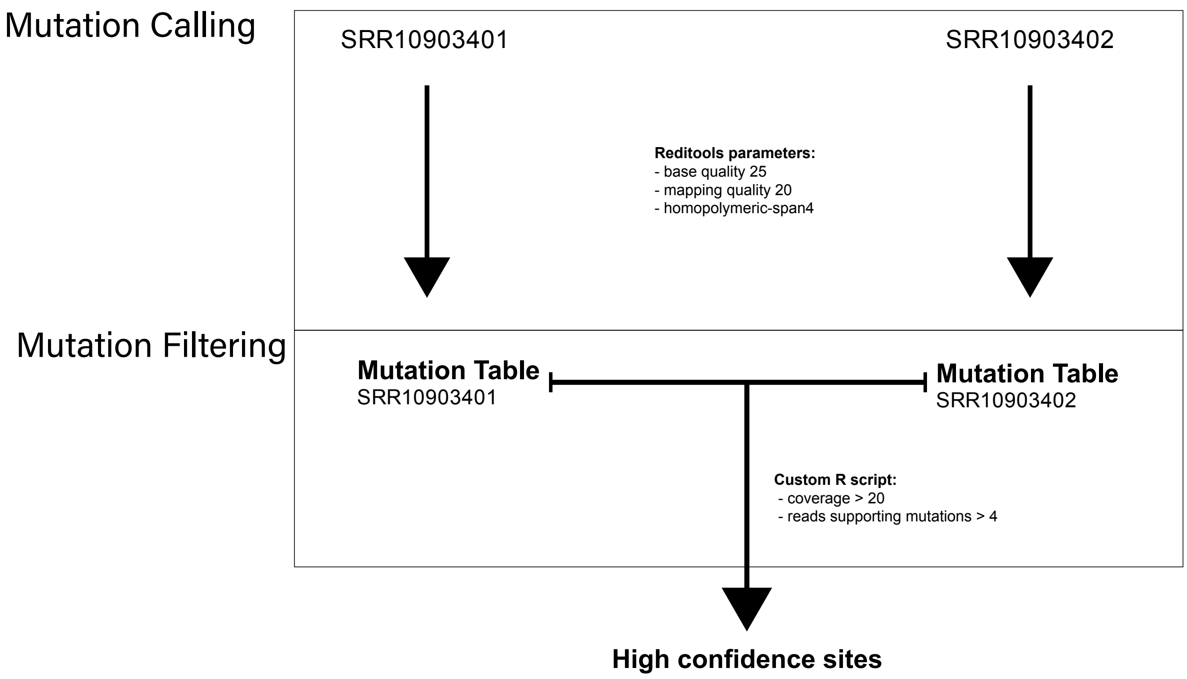


Fig. S3.

Flow-chart of the workflow. The analysis was performed in two phases: Mutation Calling and Mutation Filtering. The tools and script used in this work are described in Material and Methods.

Fig. S4

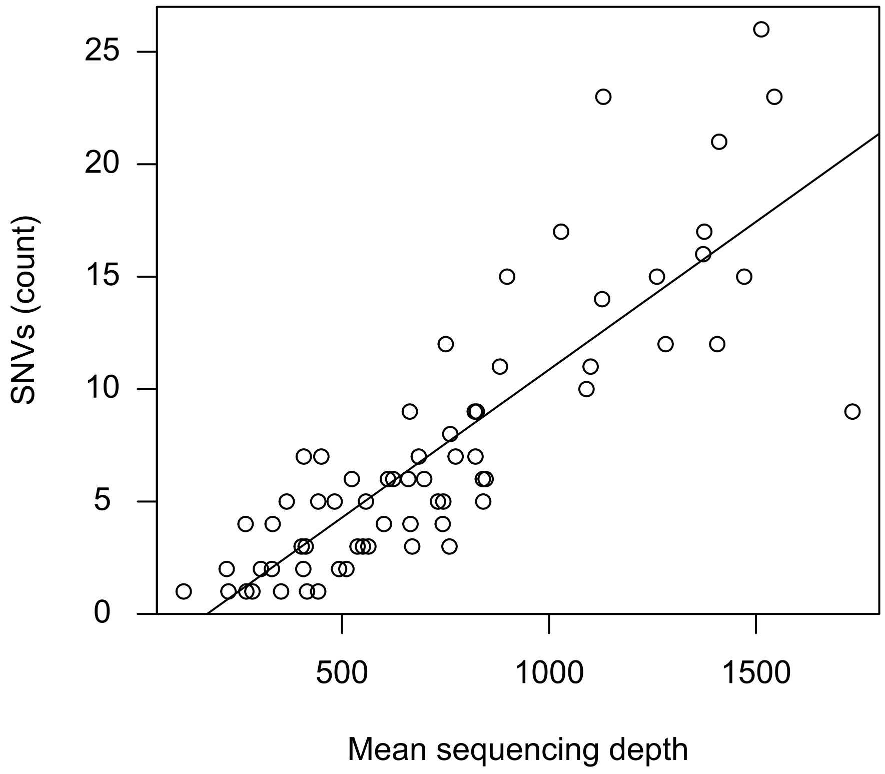


Fig. S4.

Correlation between sequencing depth and raw SNV count. Raw SNV count and mean sequencing depth were calculated using 400 nt bins. Spearman's correlation coefficient = 0.86, *p* << 0.001 .

Table S1

Run	Assay Type	BioProject	BioSample	Instrument	Host	tissue	Error Rate	Reads	Mapped Reads	% mapped read	Mean Coverage	Median
SRR10903401	RNA-Seq	PRJNA601736	SAMN13872787	Illumina MiSeq	Homo sapiens	3ronchoalveolar lavage fluic	0,17%	956612	32786	3,43%	136,96	121
SRR10903402	RNA-Seq	PRJNA601736	SAMN13872786	Illumina MiSeq	Homo sapiens	3ronchoalveolar lavage fluic	0,16%	1362751	121959	8,95%	536,59	456
SRR10971381	RNA-Seq	PRJNA603194	SAMN13922059	llumina MiniSec	Homo sapiens	Human-BALF	0,78%	56624252	260043	0,46%	603,17	413

Mean and Median are calculated on genomic range 77-29,869

Table S2.

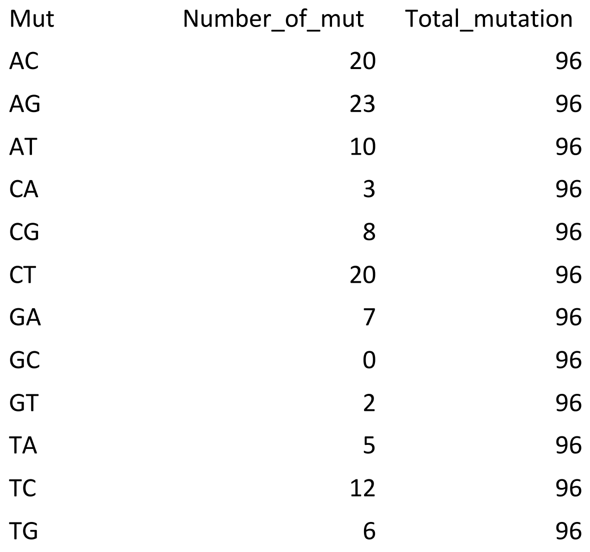
2019-nCoV genome mutations:

Mut	Number_of_mut	Total_mutation
AC	6	207
AG	28	207
AT	7	207
CA	2	207
CG	6	207
CT	65	207
GA	21	207
GC	8	207
GT	20	207
TA	15	207
TC	24	207
TG	5	207

MERS genome mutations:

Mut	Number_of_mut	Total_mutation
AC	30	1625
AG	156	1625
AT	45	1625
CA	55	1625
CG	13	1625
CT	557	1625
GA	154	1625
GC	43	1625
GT	168	1625
TA	44	1625
TC	317	1625
TG	43	1625

SARS genome mutations:



Mut	Number_of_mut	Total_mutation
AC	20	96
AG	23	96
AT	10	96
CA	3	96
CG	8	96
CT	20	96
GA	7	96
GC	0	96
GT	2	96
TA	5	96
TC	12	96
TG	6	96

References and Notes:

1. F. Wu *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* 10.1038/s41586-020-2008-3 (2020).
2. Q. Li *et al.*, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 10.1056/NEJMoa2001316 (2020).
3. E. Eisenberg, E. Y. Levanon, A-to-I RNA editing - immune protector and transcriptome diversifier. *Nat Rev Genet* **19**, 473-490 (2018).
4. M. A. O'Connell, N. M. Mannion, L. P. Keegan, The Epitranscriptome and Innate Immunity. *PLoS Genet* **11**, e1005687 (2015).
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. 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).
10. C. E. Samuel, ADARs: viruses and innate immunity. *Curr Top Microbiol Immunol* **353**, 163-195 (2012).
11. 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).
12. U. Rosani *et al.*, A-to-I editing of Malacoherpesviridae RNAs supports the antiviral role of ADAR1 in mollusks. *BMC Evol Biol* **19**, 149 (2019).
13. 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).
14. R. S. Harris *et al.*, DNA deamination mediates innate immunity to retroviral infection. *Cell* **113**, 803-809 (2003).
15. R. Mahieux *et al.*, 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).
16. C. Noguchi *et al.*, G to A hypermutation of hepatitis B virus. *Hepatology* **41**, 626-633 (2005).
17. 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).
18. R. Suspène *et al.*, 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).
19. 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).

20. A. Peretti *et al.*, 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).
21. E. N. C. Newman *et al.*, Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol* **15**, 166-170 (2005).
22. D. Pollpeter *et al.*, Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. *Nat Microbiol* **3**, 220-233 (2018).
23. A. Milewska *et al.*, APOBEC3-mediated restriction of RNA virus replication *Scientific Reports* **8**, 5960 (2018).
24. 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).
25. 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).
26. L. Perelygina *et al.*, Infectious vaccine-derived rubella viruses emerge, persist, and evolve in cutaneous granulomas of children with primary immunodeficiencies. *PLoS Pathog* **15**, e1008080 (2019).
27. L. Chen *et al.*, 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).
28. L. D. Eckerle *et al.*, Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog* **6**, e1000896 (2010).
29. L. Bazak *et al.*, 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).
30. M. Frieman *et al.*, Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *J Virol* **81**, 9812-9824 (2007).
31. 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).
32. 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 (Basel)* **10**, 13 (2018).
33. 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).
34. 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).
35. S. Osenberg, D. Dominissini, G. Rechavi, E. Eisenberg, Widespread cleavage of A-to-I hyperediting substrates. *RNA* **15**, 1632-1639 (2009).
36. 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).
37. 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).

38. 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).
39. J. Long *et al.*, A common deletion in the APOBEC3 genes and breast cancer risk. *J Natl Cancer Inst* **105**, 573-579 (2013).
40. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
41. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
42. E. Picardi, G. Pesole, REDItools: high-throughput RNA editing detection made easy *Bioinformatics* **29**, 1813-1814 (2013).
43. C. Lo Giudice, M. A. Tangaro, G. Pesole, E. Picardi, Investigating RNA editing in deep transcriptome datasets with REDItools and REDIportal. *Nat Protoc* (2020).
44. H. Pagès, P. Aboyoun, R. Gentleman, S. DebRoy, Biostrings: Efficient manipulation of biological strings. R package version 2.52.0 (2019).
45. H. Wickham, ggplot2: Elegant Graphics for Data Analysis (2016).
46. A. Mahto, splitstackshape: Stack and Reshape Datasets After Splitting Concatenated Values. R package version 1.4.8 (2019).
47. M. Morgan, H. Pagès, V. Obenchain, N. Hayden, Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 2.2.3 (2020).
48. R. C. Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing (2019).
49. 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 of 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 file S4.

Funding: The research was supported by grants from Ministero della Salute [PE-2013-02357669] and from AIRC [IG-17701];

Author contributions: Conceptualization: SDG, FM, MGT, GM, SGC; Formal Analysis, Investigation, Software: SDG, FM; Visualization: GM; Writing – original draft: GM and SGC; Writing – review&editing: SDG, FM, MGT, GM, SGC

Competing interests: Authors declare no competing interests; and

Data and materials availability: Sequencing and genomic data are available through NCBI SRA and Virus repositories, and GISAID.