

RESEARCH ARTICLE

Open Access



# SARS-CoV-2 and human retroelements: a case for molecular mimicry?

Benjamin Florian Koch\*

## Abstract

**Background:** The factors driving the late phase of COVID-19 are still poorly understood. However, autoimmunity is an evolving theme in COVID-19's pathogenesis. Additionally, deregulation of human retroelements (RE) is found in many viral infections, and has also been reported in COVID-19.

**Results:** Unexpectedly, coronaviruses (CoV) – including SARS-CoV-2 – harbour many RE-identical sequences (up to 35 base pairs), and some of these sequences are part of SARS-CoV-2 epitopes associated to COVID-19 severity. Furthermore, RE are expressed in healthy controls and human cells and become deregulated after SARS-CoV-2 infection, showing mainly changes in long interspersed nuclear element (LINE1) expression, but also in endogenous retroviruses.

**Conclusion:** CoV and human RE share coding sequences, which are targeted by antibodies in COVID-19 and thus could induce an autoimmune loop by molecular mimicry.

**Keywords:** Coronaviruses, SARS-CoV-2, COVID-19 epitope signatures, Autoimmunity, Molecular mimicry, Human retroelements, Long interspersed nuclear elements (LINE), Endogenous retroviruses (ERV)

## Background

At the end of 2019, a severe acute respiratory syndrome (SARS)-like disease was noted in eastern China and a novel coronavirus (later designated SARS-CoV-2) recognized as the factor for the disease, COVID-19 [1]. By the spring of 2022, 447 million people have been infected globally, with 6 million casualties [2]. COVID-19 can be divided into an early viral replication phase and a late stage of organ failure [3, 4]. While the inhibition of SARS-CoV-2 replication has already been achieved [5–10], the factors driving the late phase of the disease are poorly understood [11, 12]. However, it has been reported that autoimmunity [13–27] and deregulation of human retroelements (RE) might contribute to the outcome of COVID-19 patients [28–31].

The RE share a reverse transcriptase as a common denominator. Together with an endonuclease, they can move by “copy and paste.” Based on the presence of an envelope gene, they can be divided into long terminal repeat (LTR) positive and LTR negative retrotransposons. The former and endogenous retroviruses (ERV) belong to LTR positive elements. Long interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE) and SVA elements (SINE-R, VNTR and Alu) belong to LTR negative elements [32–35]. The LINE contain at least two open reading frames (ORFs), ORF1, coding for a nucleic acid binding protein with chaperone activity (ORF1p) and ORF2, which codes for a reverse transcriptase/endonuclease (ORF2p) [35, 36]. Importantly, RE make up 50 – 70% of the human genome [37, 38]. About 20% of the genome is made up from LINE sequences (c. 500,000 copies), of which more than 100 LINE1 family members are still intact and about 68 active in humans. The LINE1 show strong interpersonal differences [39, 40] and an age-dependent expression pattern

\*Correspondence: B.Koch@med.uni-frankfurt.de  
Department of Internal Medicine, Nephrology, Goethe University Hospital, Johann Wolfgang Goethe University Frankfurt/Main, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

[41–43]. By comparison, ERV make up about 8% of the human genome. Despite – similar to LINE – predominant inactivation, there are still hundreds of intact viral promoters and open reading frames from which the expression of ERV transcripts and proteins is possible [44–46]. The RE activation is known from many viral infections, such as HIV [47], dengue [48], influenza A [48], Zika virus [48], West Nile virus [48], measles [48], Epstein-Barr virus [49] and cytomegalovirus [50]. Therefore, I looked for the relationship of coronaviruses (CoV) to human RE based on genome, transcriptome, epitope and peptide array data. Here, transcriptome analysis coincidentally revealed many RE-identical sequences and shared epitopes in the CoV family members investigated, such as SARS-CoV-2, MERS-CoV and HKU1. To the best of my knowledge, these findings have never been reported. Importantly, epitopes are shared between human LINE1- and SARS-CoV-2 proteins and antibodies against some of these epitopes have been found to be correlated to COVID-19’s severity. In addition, RE are

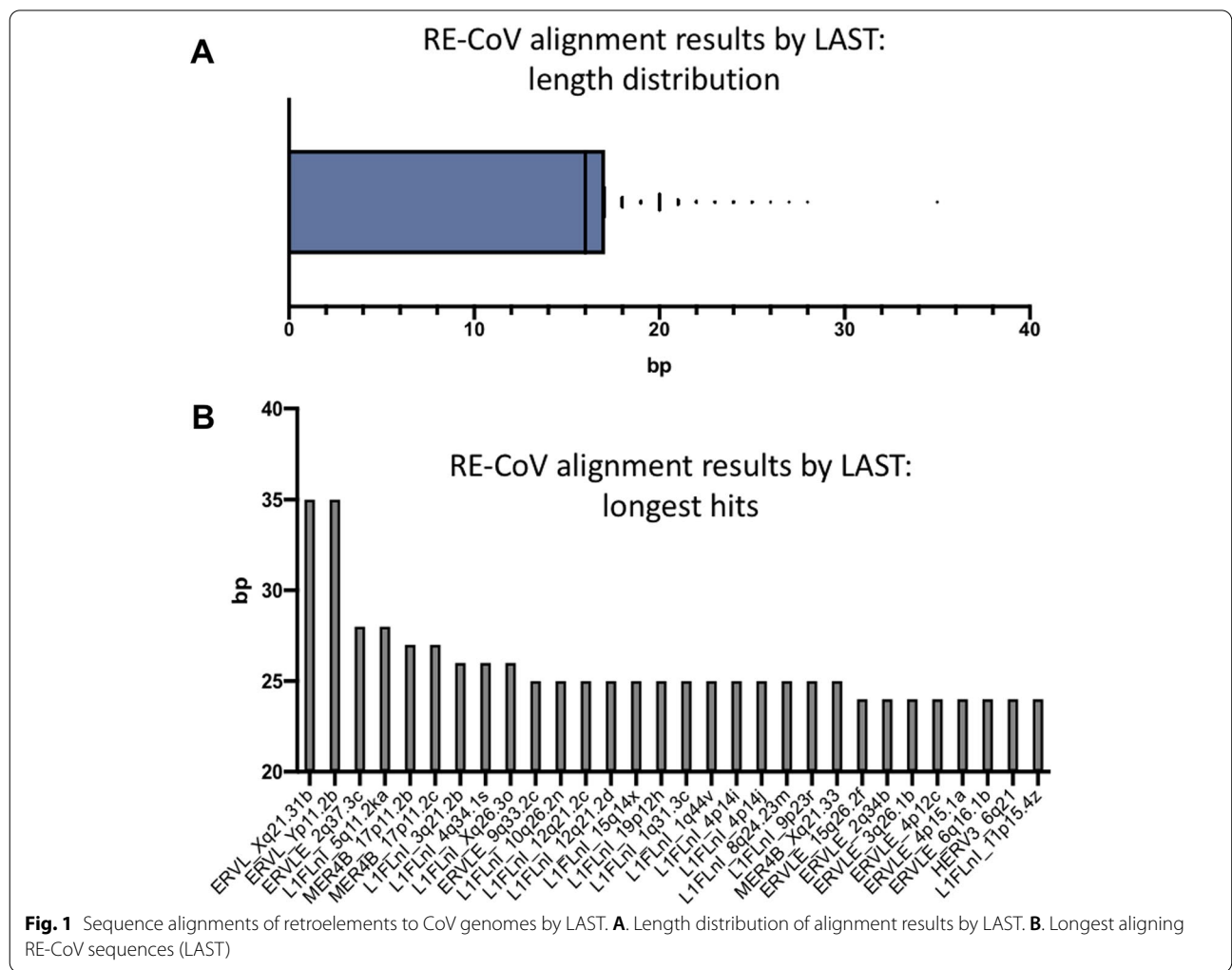
expressed in healthy controls and deregulated in COVID-19 patients, as well as in SARS-CoV-2-infected human cells.

**Results**

The CoV genomes harbour a large number of RE-identical sequences. Several of these sequences represent shared RE-SARS-CoV-2 epitopes. Importantly, antibodies against some of these epitopes are correlated to the severity of COVID-19. In addition, RE are widely expressed in healthy controls and deregulated in COVID-19 patients, as well as in SARS-CoV-2-infected human cells.

**Sequence identity between retroelements and coronaviruses**

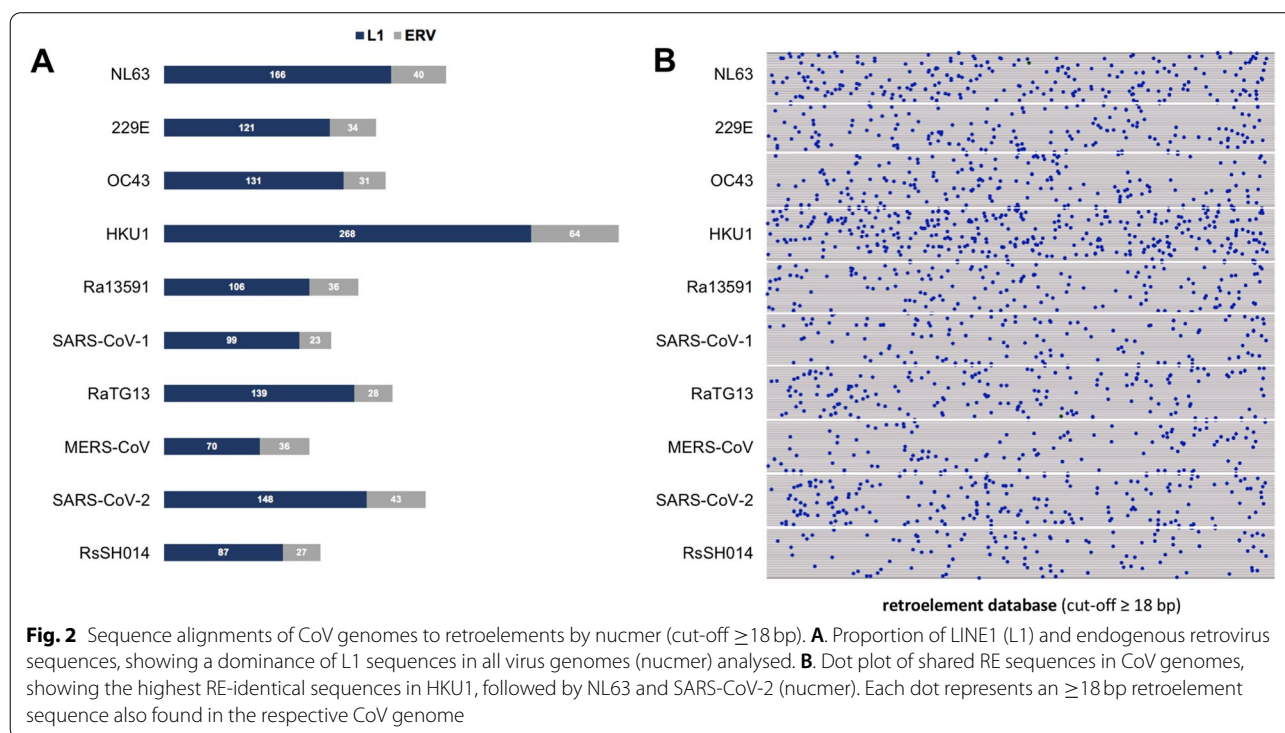
A sequence identity ( $\geq 12$  bp, range 12 – 35 bp, Fig. 1A) of human RE sequences to CoV genomes from SARS-CoV-2, SARS-CoV-1, MERS-CoV, NL63, 229E, OC43, HKU1, bat CoV RA13591, bat CoV RATG13 and bat CoV



RSSH014 was found by sequence alignment of human RE sequences and different CoV genomes (Figs. 1 and 2, Table 1). Very high counts of RE-identical sequences in CoV were seen at  $\geq 12$ ,  $\geq 15$  and  $\geq 18$  bp (Table 1).

A cut-off  $\geq 18$  bp (correlating to potential epitopes of at least 6 aa) was chosen for downstream analysis for sensitivity and epitope size reasons. A 6 aa cut-off corresponds well to a known immuno-relevant linear epitope length of 4 – 12 aa, as about 50% of them have a length  $\leq 8$  aa (about 25%  $\leq 6$  aa, and only a few of 4 aa) [51]. At this

cut-off point, the majority of RE-identical sequences are seen in HKU1 (332), followed by NL63 (206) and SARS-CoV-2 (191) (Fig. 2A and B, Table 1). SARS-CoV-2 and RE sequence data were further explored by “LAST” in order to allow single nucleotide polymorphisms to be included, thereby alignments to RE sequences up to 35 bp were seen (Supplementary Table 2). In the RE-CoV data, LINE1 represent the majority of all shared sequences, while alignment to ERV sequences is a relevant minority and includes the 35 bp hits (Fig. 1B, Supplementary



**Table 1** Number of retroelement-identical sequences in CoV genomes dependent on sequence length (12 – 27 bp, based on 100% sequence identity (alignment by nucmer). Underlines indicate the highest score at the respective cut-off

CoV	Genbank ID	RE-identical sequences (nucmer)					
		$\geq 12$ bp	$\geq 15$ bp	$\geq 18$ bp	$\geq 21$ bp	$\geq 24$ bp	$\geq 27$ bp
229E	NC_002645.1	16,992	4663	<b>155</b>	8	5	4
NL63	NC_005831.2	17,751	5758	<b>206</b>	4	0	0
OC43	AY391777.1	<u>19,428</u>	5709	<b>162</b>	3	0	0
HKU1	NC_006577.2	19,112	<u>6843</u>	<b>332</b>	<u>13</u>	0	0
MERS-CoV	NC_019843.3	18,435	4252	<b>106</b>	3	1	0
SARS-CoV-1	AY291315.1	18,446	4731	<b>122</b>	2	1	0
SARS-CoV-2	NC_045512.2	18,917	5358	<b>191</b>	11	<u>3</u>	2
RsSH014	KC881005.1	18,425	4651	<b>114</b>	5	1	0
Ra13591	MG916904.1	17,971	4644	<b>142</b>	2	0	0
RaTG13	MN996532.2	18,950	5327	<b>167</b>	4	0	0

Tables 1 and 2). In conclusion, genome analysis revealed the presence of many short RE-identical sequences in CoV genomes, including SARS-CoV-2.

**Shared epitopes between SARS-CoV-2- and retroelement proteins**

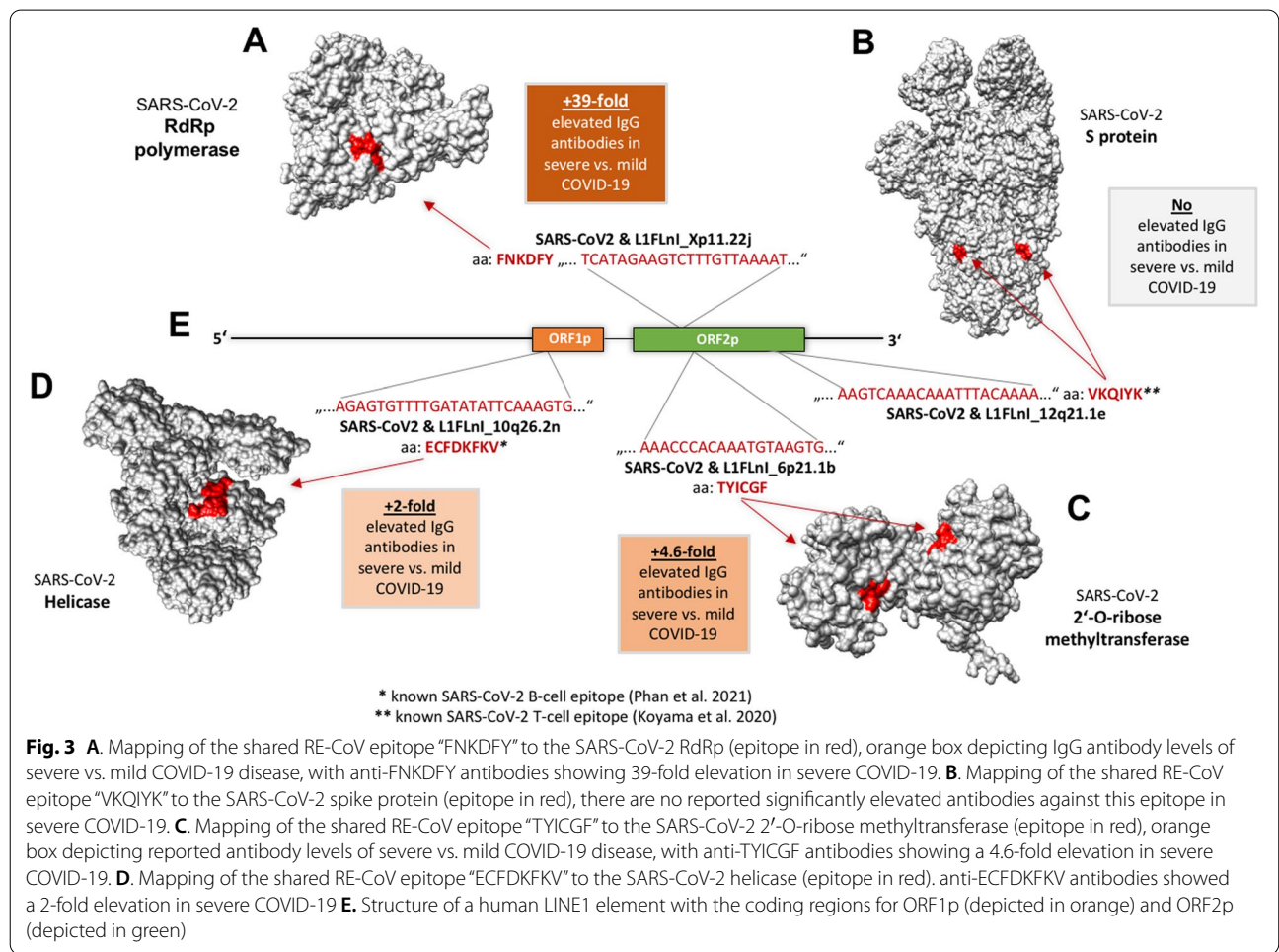
Subsequently, all RE-identical sequences  $\geq 18$ bp were compared to the coding regions of the genome of SARS-CoV-2. Accordingly, 70 sequences showing identical aa sequences in CoV and RE were identified (Supplementary Table 1). These sequences were then compared to results from a peptide array, which investigated epitope signatures in COVID-19 patients (severe vs. mild) [52]. An overlap of human LINE1 proteins to SARS-CoV-2 epitopes from the RNA-dependent RNA polymerase (RdRp), helicase and 2'-O-ribose methyltransferase was detected for epitopes targeted with  $>2$ -fold elevated antibody levels in severe cases (Fig. 3). Importantly, antibodies targeting an epitope of the SARS-CoV-2 RdRp polymerase, which is identical to an epitope of the LINE1 ORF2p endonuclease domain, were 39-fold elevated in

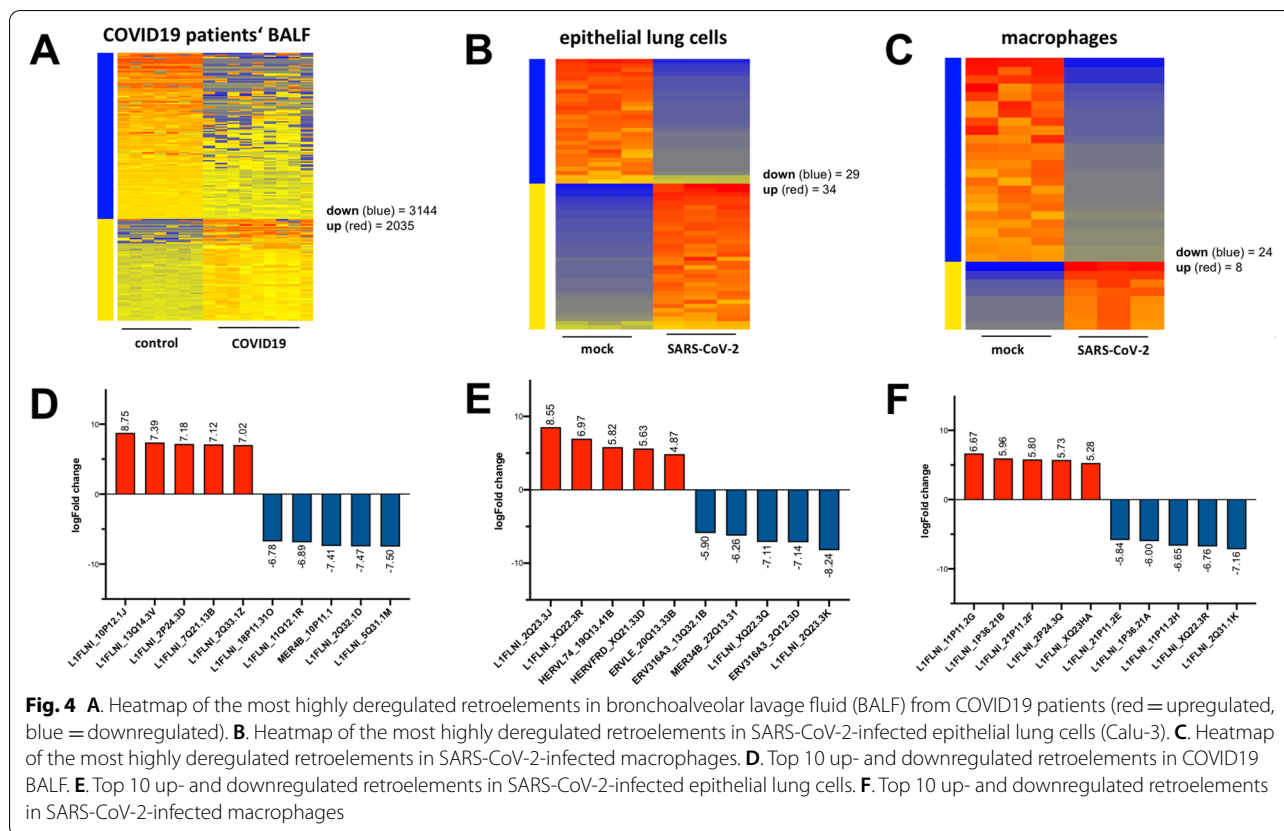
severely compared to only mildly affected COVID-19 patients (Fig. 3A). The same is seen with antibodies targeting the shared CoV-RE epitopes from the 2'-O-ribose methyltransferase (Fig. 3C) and helicase (Fig. 3D). The latter is also a known B cell epitope, aa "PARARVECFDKFKV" (the known B cell epitope is depicted in bold) [53]. Many other shared RE-CoV peptides (similar to those displayed in Fig. 3B) were not targeted by antibodies in severe vs. mild COVID-19 (Supplementary Table 2), but some are known as T cell epitopes, such as the one present in all three chains of the spike protein shown in Fig. 3B (aa **VKQIYK**TPPIKDF, the known T cell epitope sequence is depicted in bold) [54].

Taken together, SARS-CoV-2 and RE share peptide sequences, of which some are epitopes correlated to COVID-19 severity.

**Transcriptome analysis of retroelements in SARS-CoV-2-infected cells**

An RE analysis of COVID-19 patient data (bronchoalveolar lavage fluid, BALF), SARS-CoV-2 infected lung





epithelial cells and SARS-CoV-2 infected macrophages was performed to explore the presence of and changes in RE expression after SARS-CoV-2 infection. Infection resulted in a highly significant (adjusted  $p$ -value  $\leq 0.05$ ) and relevant (fold change  $\geq 2$ ) deregulation of human RE in all samples. Transcriptome data from COVID-19 patients' BALF compared to healthy controls shows an upregulation of 2035 and downregulation of 3144 RE (Fig. 4A). Among the top deregulated RE are mainly LINE1 (Fig. 4D). SARS-CoV-2-infected epithelial lung cells (Calu-3) show 34 up- and 29 downregulated RE (Fig. 4E), while infected human macrophages have 8 up- and 24 downregulated RE. Among the top de-regulated RE for both are also mainly LINE1 (Fig. 4E, F).

In conclusion, RE are expressed in COVID-19 patients and human cells and become deregulated after SARS-CoV-2 infection, showing mainly changes in LINE1 expression.

### Discussion

The factors driving the late phase of COVID-19 are still not fully understood [11, 12]. However, there is evidence that autoantibodies and autoreactive lymphocytes could contribute to the disease's final outcome [13–27]. Therefore, the question of autoantibody formation in

COVID-19 has to be asked. The employment of a comprehensive RE database revealed many RE-identical sequences in ten CoV family members investigated, such as in SARS-CoV-2, MERS-CoV and HKU1 (Figs. 1 and 2). Crucially, it was found that the LINE1 proteins ORF1p and ORF2p have peptides identical to SARS-CoV-2 epitopes (Fig. 3), and that some of these epitopes are associated with COVID-19's severity, as shown by correlation to COVID-19 patients' antibody titres (Fig. 3). In addition, RE are deregulated in COVID-19 patients (Fig. 4A), as well as SARS-CoV-2-infected human epithelial lung cells and macrophages (Fig. 4B and C), which has occasionally been reported in the last few months for cell lines and patients [28–31]. Among the analysed REs, LINE1 are strongly represented in all results (Figs. 2, 3 and 4, Supplementary Table 1 and 2). The LINE1 code for at least a nucleic acid binding protein with chaperone activity (ORF1p) and a reverse transcriptase/endonuclease (ORF2p). Importantly, autoantibodies targeting the LINE1 ORF2p endonuclease domain have been reported in 41% of SARS-CoV-1 patients [55]. The RE are also targeted by autoantibodies in several connective tissue diseases, for example, antibodies against LINE1's ORF1p or ERV HERV-K's envelope protein have been described in patients with systemic lupus erythematosus, lupus

nephritis, rheumatoid arthritis, Sjogren's syndrome and mixed connective tissue disease [56–65]. Relating to SARS, the autoantibodies' target, LINE1 ORF2p, was prominently stained post-mortem in lung macrophages (residing in blood vessels), leading the authors to suspect a build-up of autoreactive CD4<sup>+</sup> Th cells and, thus, an autoimmune loop in SARS [55]. Importantly, there is also increasing evidence for an autoimmune pathogenesis in severe COVID-19 [13–27, 66, 67]. One explanation for autoantibody formation is by molecular mimicry, i.e. shared epitopes between pathogens and hosts [68–72]. The evolution of mimicry epitopes in pathogens could be based on chance. However, although the RE-identical sequences in CoV observed are short (12 – 35 bp), the sequence lengths observed make formation by chance highly unlikely. Exemplarily, taking the genetic code (A, T, C, G) raised to a sequence of 18 bp ( $4^{18}$ ) results in 68,719,476,736 possible bp combinations, thus, the chance of getting one identical sequence is 1:69 billion. Additionally, a myriad of 12 bp events (Table 1) occurring by chance is stochastically very unlikely ( $4^{12} = 16,777,216$ ) at more than 18,000 events. Moreover, an observed 35 bp hit such as ERVL\_Xq21.31b ( $4^{35}$ ) corresponds to 1.18 E21 possible bp combinations, thus, the chance of getting an identical sequence is 1:1.1 trillion – without accounting for all the other matching sequences. Therefore, recombination activities more probably account for the phenomena observed. The exchange of genetic material by recombination in RNA viruses is generally associated with virulence, host range and host response [73]. It is known that recombination in CoV can take place during co-infections at a high frequency by homologous and non-homologous recombination [74–76]. Mechanistically, an explanation could be the switching of the RdRp between multiple available RNA strands during replication [77]. This could have happened in a CoV host/ancestor with relevant LINE1 expression, as this is possible in some bat species. The black-bearded tomb bat (*Taphozous melanopogon*), for example, harbours two active LINE families [78] and shows relevant SARS-CoV-2 infection efficiency [79]. Moreover, lots of ERV families also reside in bats [80]. Therefore, serial acquisition of RE sequences, possibly taken from CoV in host animals (starting many million years ago) is a feasible scenario. Relating to the rather short sequence lengths observed, there might be an evolutionary functional constraint working against the uptake of longer RE sequences, but a benefit for the virus by coating itself with host self-antigens (“self-peptide coat”). This would dampen the innate and adaptive immune response by the presentation of “viral but self-like” peptides. The consequence of this hypothesis is in line with the view of autoimmune disease as a breakdown of self-tolerance

[81, 82]. Based on the findings, autoantibodies targeting human RE could be a factor in CoV-induced disease, like COVID-19. However, this report has limitations, as the data basis for a more extensive analysis of anti-RE autoantibodies in COVID-19 still does not exist.

## Conclusion

In conclusion, it was found that CoV – including SARS-CoV-2 – harbour many RE-identical sequences, and that some of these sequences are part of SARS-CoV-2 epitopes associated with COVID-19 severity.

## Methods

### Genome analysis

Genome sequences from SARS-CoV-2 (isolate NC045512.2=Wuhan-Hu-1), SARS-CoV-1 (AY291315.1=FFM1), MERS-CoV (NC\_019843.3=EMC2012), human pathogenic CoVs (NC-006577.2=HKU1; AY391777.1=OC43, NC-002645.1=229E; NC-005831.2=NL63) and bat CoVs (MN996532.2=RaTG13, KC881005.1=RsSHC014; MG916904.1=Ra1359) were downloaded from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>). Retro.hg38.v1 ([https://github.com/mlbendall/telescope\\_annotation\\_db/tree/master/builds](https://github.com/mlbendall/telescope_annotation_db/tree/master/builds)) was employed as an RE database. The database contains 28.513 RE and is made of “RepeatMasker” hits for 60 HERV families (RepeatMasker Open-4.0, <http://www.repeatmasker.org/>) and all LINE elements from “L1base v2” (<https://l1base.charite.de/>) [83]. Alignment of the retro.hg38.v1 database to CoV genomes was done by the genome sequence aligner “nucmer” [84] (4.0.0beta2) on [galaxy.org](https://galaxy.org) [85] and a local installation of “LAST” (v1250), a programme for genome scale sequence comparison [86]. The minimum sequence length cut-off (with 100% sequence identity) was stepwise chosen at 12, 15, 18, 21, 24, and  $\geq 27$ , based on an immuno-relevant epitope size of about 4 – 12 amino acids (aa) (many epitopes are less than 8 aa, about 25%  $\leq 6$  aa, but only a few at 4 aa [51]). The nucmer “-b” and “-L” variables were used accordingly, and “Show-Coords” as well as “Mummerplot” from the “MUMmer 4” package [84] were employed to extract and plot data. Regarding to “LAST”, firstly, an RE database was built (“lastdb -uNEAR -c RE\_db retro.hg38.v1.fa”) and then CoV genomes were compared to the RE database (“lстал -D100 RE\_db CoV\_genome.fa > RE\_db\_CoV.maf”).

### Epitope-specific antibody data in COVID-19 patients

The SARS-CoV-2 epitope-specific antibody data (IgG) in severely vs. mildly affected COVID-19 patients are from Schwarz et al. [52] “Peptide microarray data – severe vs. mild – IgG,” with the peptides: 1060 (NSP12, QTVK-PGNFNKDFYDF, LogFC 5.3, *p*-value 2.4E-04, FDR-adj. *p*-value 2.8E-02), 1243 (NSP16, ENDSKEGFFTYICGF, LogFC 2.2, *p*-value 4.0E-02, FDR-adj. *p*-value 5.2E-01), 1227 (NSP13, IPARARVECFDKFKV, LogFC –0.9,

p-value 3.2E-01, FDR-adj. p-value 5.3E-01) and 1690 (Spike, AQVKQIYKTPPIKDF, LogFC 0.2, p-value 8.3E-01, FDR-adj. p-value 8.5E-01). “L1base v2” was used for comparison with coding LINE1 sequences (<https://l1base.charite.de/>) [83]. Known SARS-CoV-2 B- and T-cell epitopes are from Phan et al. [53] and Griffoni et al. [54]. The PDB data for the SARS-CoV-2 RdRp (PDB ID: 7BW4), helicase (PDB ID: 7NNG), 2'-O-ribose methyltransferase (PDB ID: 7JYY) and -spike protein (PDB ID: 7LSS) were downloaded from <https://www.rcsb.org> and epitopes displayed by “UCSF Chimera v1.15” (for Mac OS) [87].

### Transcriptome analysis

Total RNA sequencing data from SARS-CoV-2-infected macrophages (BioProject ID PRJNA637580, Sequence Read Archive (SRA) ID mock: SRR11934391, SRR11934392, SRR11934393, infected: SRR11934394, SRR11934395, SRR11934396) [88], Calu-3 adenocarcinomic lung epithelial cells (PRJNA615032, mock: SRR11517744, SRR11517745, SRR11517746, infected: SRR11517747, SRR11517748, SRR11517749) [89] and bronchoalveolar lavage (BALF) samples from intensive care COVID-19 patients (PRJNA605983SRA, SRA: SRR11092056, SRR11092057, SRR11092058, SRR11092059, SRR11092060, SRR11092061, SRR11092062, SRR11092063, SRR11092064) [90] compared to healthy controls (PRJNA316136, SRA: SRR3286988, SRR3286989, SRR3286990, SRR3286991, SRR5515942, SRR5515943, SRR5515944) [91] were downloaded from SRA (<https://www.ncbi.nlm.nih.gov/sra>), quality controlled by FastQC (Babraham Institute, Cambridge, UK, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and Illumina adapters trimmed by Trimmomatic [92]. Salmon [93] and DESeq2 [94] were employed for differential RE analysis, with standard parameters after indexing the retro.hg38.v1 database (“salmon index -t retro.hg38.v1.fa -i retro.hg38.v1\_index -k 31”). Heatmaps were done by iDEP v0.92 [95] and graphs by GraphPad Prism software version 8.0 for OS X (GraphPad Software Inc., USA).

### Abbreviations

229E: Human coronavirus 229E; aa: Amino acids; BALF: Bronchoalveolar lavage; bp: Base pairs; CoV: Coronavirus; COVID-19: Coronavirus disease 2019; ERV: Endogenous retroviruses; HKU1: Human coronavirus HKU1; LINE1: Long interspersed nuclear elements; LTR: Long terminal repeat; MERS-CoV: Middle East respiratory syndrome-related coronavirus; NL63: Human coronavirus NL63; OC43: Human coronavirus OC43; ORF: Open reading frame; Ra1359: Bat coronavirus Ra1359; RaTG13: Bat coronavirus RaTG13; RdRp: RNA-dependent RNA polymerase; RE: Human retroelements; RsSHC014: Bat coronavirus RsSHC014; SARS-CoV-1: Severe acute respiratory syndrome coronavirus type 1; SARS-CoV-2: Severe acute respiratory syndrome coronavirus type 2; SINE: Short interspersed nuclear elements; SRA: Sequence read archive; SVA: SINE-R, VNTR and Alu.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12863-022-01040-2>.

**Additional file 1: Supplementary Table 1.** RE – CoV sequence alignment results by nucmer.

**Additional file 2: Supplementary Table 2.** RE – CoV sequence alignment results by LAST.

### Acknowledgements

The author gratefully acknowledges Philip Saunders for valuable comments and proof-reading.

### Author's contributions

Conceptualization, investigation, formal analysis, writing: BK. The author read and approved the final manuscript.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Availability of data and materials

All data generated in this study are included in this article and its supplementary files. The data used in this study are openly available at the sources detailed in the methods section.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The author declares no competing interests.

Received: 23 November 2021 Accepted: 16 March 2022

Published online: 08 April 2022

### References

- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *New Engl J Med*. 2020;382:727–33.
- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020. [https://doi.org/10.1016/s1473-3099\(20\)30120-1](https://doi.org/10.1016/s1473-3099(20)30120-1).
- Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, transmission, and pathogenesis of SARS-CoV-2. *BMJ*. 2020;371:m3862.
- Khourssaji M, Chapelle V, Evenepoel A, Belkhir L, Yombi JC, van Dievoet MA, et al. A biological profile for diagnosis and outcome of COVID-19 patients. *Clin Chem Lab Med*. 2020;58:2141–50.
- Ellinger B, Bojkova D, Zaliani A, Cinatl J, Claussen C, Westhaus S, et al. A SARS-CoV-2 cytopathicity dataset generated by high-content screening of a large drug repurposing collection. *Sci Data*. 2021;8:70.
- Bojkova D, Bechtel M, McLaughlin K-M, McGreig JE, Klann K, Bellinghausen C, et al. Aprotinin inhibits SARS-CoV-2 replication. *Cells*. 2020;9:2377.
- Klann K, Bojkova D, Tascher G, Ciesek S, Münch C, Cinatl J. Growth factor receptor signaling inhibition prevents SARS-CoV-2 replication. *Mol Cell*. 2020;80:164–74 e4.
- Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature*. 2020;583:469–72.
- Jeon S, Ko M, Lee J, Choi I, Byun SY, Park S, et al. Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs. *Antimicrob Agents Ch*. 2020;64:e00819–20.

10. Mostafa A, Kandeil A, Elshaier YAMM, Kutkat O, Moatasim Y, Rashad AA, et al. FDA-approved drugs with potent in vitro antiviral activity against severe acute respiratory syndrome coronavirus 2. *Pharmaceuticals (Basel)*. 2020;13:443.
11. Satturwar S, Fowkes M, Farver C, Wilson AM, Eccher A, Girolami I, et al. Postmortem findings associated with SARS-CoV-2. *Am J Surg Pathol*. 2021;45:587–603.
12. Menter T, Haslbauer JD, Nienhold R, Savic S, Hopfer H, Deigendesch N, et al. Postmortem examination of COVID-19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings in lungs and other organs suggesting vascular dysfunction. *Histopathology*. 2020;77:198–209.
13. Khamsi R. Rogue antibodies could be driving severe COVID-19. *Nature*. 2021;590:29–31.
14. Bastard P, Rosen LB, Zhang Q, Michailidis E, Hoffmann H-H, Zhang Y, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science*. 2020;370:eabd4585.
15. Wang EY, Mao T, Klein J, Dai Y, Huck JD, Jaycox JR, et al. Diverse functional autoantibodies in patients with COVID-19. *Nature*. 2021;595:283–8.
16. Icenogle T. COVID-19: infection or autoimmunity. *Front Immunol*. 2020;11:2055.
17. Zuniga M, Gomes C, Carsons SE, Bender MT, Cotzia P, Miao QR, et al. Autoimmunity to the lung protective phospholipid-binding protein Annexin A2 predicts mortality among hospitalized COVID-19 patients. *Eur Respir J*. 2021;58:2100918.
18. Ehrenfeld M, Tincani A, Andreoli L, Cattalini M, Greenbaum A, Kanduc D, et al. Covid-19 and autoimmunity. *Autoimmun Rev*. 2020;19:102597.
19. Zuo Y, Estes SK, Ali RA, Gandhi AA, Yalavarthi S, Shi H, et al. Prothrombotic autoantibodies in serum from patients hospitalized with COVID-19. *Sci Transl Med*. 2020;12:eabd3876.
20. Obando-Pereda G. Can molecular mimicry explain the cytokine storm of SARS-CoV-2?: an in silico approach. *J Med Virol*. 2021. <https://doi.org/10.1002/jmv.27040>.
21. Lucchese G, Flöel A. Guillain-Barré syndrome, SARS-CoV-2 and molecular mimicry. *Brain*. 2021;144:e43. <https://doi.org/10.1093/brain/awab067>.
22. Franke C, Ferse C, Kreye J, Reincke SM, Sanchez-Sendin E, Rocco A, et al. High frequency of cerebrospinal fluid autoantibodies in COVID-19 patients with neurological symptoms. *Brain Behav Immun*. 2020;93:415–9.
23. Vojdani A, Vojdani E, Kharrazian D. Reaction of human monoclonal antibodies to SARS-CoV-2 proteins with tissue antigens: implications for autoimmune diseases. *Front Immunol*. 2021;11:617089.
24. Novelli L, Motta F, Santis MD, Ansari AA, Gershwin ME, Selmi C. The JANUS of chronic inflammatory and autoimmune diseases onset during COVID-19 – a systematic review of the literature. *J Autoimmun*. 2020;117:102592.
25. Saini SK, Hersby DS, Tamhane T, Povlsen HR, Hernandez SPA, Nielsen M, et al. SARS-CoV-2 genome-wide T cell epitope mapping reveals immunodominance and substantial CD8+ T cell activation in COVID-19 patients. *Sci Immunol*. 2021;6:eabf7550.
26. Gauchotte G, Venard V, Segondy M, Cadoz C, Esposito-Fava A, Barraud D, et al. SARS-Cov-2 fulminant myocarditis: an autopsy and histopathological case study. *Int J Legal Med*. 2021;135:577–81.
27. Lagadinou M, Zareifopoulos N, Gkentzi D, Sampsonas F, Kostopoulou E, Marangos M, et al. Alterations in lymphocyte subsets and monocytes in patients diagnosed with SARS-CoV-2 pneumonia: a mini review of the literature. *Eur Rev Med Pharmacol*. 2021;25:5057–62.
28. Balestrieri E, Minutolo A, Petrone V, Fanelli M, Iannetta M, Malagnino V, et al. Evidence of the pathogenic HERV-W envelope expression in T lymphocytes in association with the respiratory outcome of COVID-19 patients. *Ebiomedicine*. 2021;66:103341.
29. El-Shehawi AM, Alotaibi SS, Elseehy MM. Genomic study of COVID-19 Corona virus excludes its origin from recombination or characterized biological sources and suggests a role for HERVs in its wide range symptoms. *Cytol Genet*. 2020;54:588–604.
30. Souza T, Temerozo J, Fintelman-Rodrigues N, Santos MC, Hottz E, Sacramento C, et al. Human endogenous retrovirus K activation in the lower respiratory tract of severe COVID-19 patients associates with early mortality; 2021. <https://doi.org/10.21203/rs.3.rs-514541/v1>.
31. Li M, Schifanella L, Larsen PA. Alu retrotransposons and COVID-19 susceptibility and morbidity. *Hum Genomics*. 2021;15:2.
32. McDonald TL, Zhou W, Castro CP, Mumm C, Switzenberg JA, Mills RE, et al. Cas9 targeted enrichment of mobile elements using nanopore sequencing. *Nat Commun*. 2021;12:3586.
33. Marshall JN, Lopez AI, Pfaff AL, Koks S, Quinn JP, Bubb VJ. Variable number tandem repeats – their emerging role in sickness and health. *Exp Biol Med*. 2021;246(12):1368–76. <https://doi.org/10.1177/15353702211003511>.
34. Deininger PL, Batzer MA. Mammalian Retroelements. *Genome Res*. 2002;12:1455–65.
35. Mangiacavalli A, Liu P, Valle FD, Orlando V. New insights into the functional role of retrotransposon dynamics in mammalian somatic cells. *Cell Mol Life Sci*. 2021;78:5245–56.
36. Carnell AN, Goodman JL. The long (LINEs) and the short (SINEs) of it: altered methylation as a precursor to toxicity. *Toxicol Sci*. 2003;75:229–35.
37. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
38. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet*. 2011;7:e1002384.
39. Strevva VA, Jordan VE, Linker S, Hedges DJ, Batzer MA, Deininger PL. Sequencing, identification and mapping of primed L1 elements (SIMPLE) reveals significant variation in full length L1 elements between individuals. *BMC Genomics*. 2015;16:220.
40. Martin SL. On the move. *Elife*. 2018;7:e34901.
41. Simon M, Meter MV, Ablaeva J, Ke Z, Gonzalez RS, Taguchi T, et al. LINE1 Derepression in aged wild-type and SIRT6-deficient mice drives inflammation. *Cell Metab*. 2019;29:871–85 e5.
42. Mahmood W, Erichsen L, Ott P, Schulz WA, Fischer JC, Arauzo-Bravo MJ, et al. Aging-associated distinctive DNA methylation changes of LINE-1 retrotransposons in pure cell-free DNA from human blood. *Sci Rep*. 2020;10:22127.
43. Roberson PA, Romero MA, Osburn SC, Mumford PW, Vann CG, Fox CD, et al. Skeletal muscle LINE-1 ORF1 mRNA is higher in older humans but decreases with endurance exercise and is negatively associated with higher physical activity. *J Appl Physiol*. 2019;127:895–904.
44. Villesen P, Aagaard L, Wiuf C, Pedersen FS. Identification of endogenous retroviral reading frames in the human genome. *Retrovirology*. 2004;1:32.
45. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet*. 2012;13:283–96.
46. Hohn O, Hanke K, Bannert N. HERV-K(HML-2), the best preserved family of HERVs: Endogenization, expression, and implications in health and disease. *Front Oncol*. 2013;3:246.
47. Vincendeau M, Götttesdorfer I, Schreml JMH, Wetie AGN, Mayer J, Greenwood AD, et al. Modulation of human endogenous retrovirus (HERV) transcription during persistent and de novo HIV-1 infection. *Retrovirology*. 2015;12:27.
48. Wang M, Wang L, Liu H, Chen J, Liu D. Transcriptome analyses implicate endogenous retroviruses involved in the host antiviral immune system through the interferon pathway. *Virology*. 2021;36:1315–26.
49. Sutkowski N, Conrad B, Thorley-Lawson DA, Huber BT. Epstein-Barr virus Transactivates the human endogenous retrovirus HERV-K18 that encodes a Superantigen. *Immunity*. 2001;15:579–89.
50. Assinger A, Yaiw K-C, Götttesdorfer I, Leib-Mösch C, Söderberg-Nauclér C. Human cytomegalovirus (HCMV) induces human endogenous retrovirus (HERV) transcription. - PubMed - NCBI. *Retrovirology*. 2013;10:132.
51. Buus S, Rockberg J, Forsström B, Nilsson P, Uhlen M, Schafer-Nielsen C. High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays\*. *Mol Cell Proteomics*. 2012;11:1790–800.
52. Schwarz T, Heiss K, Mahendran Y, Casilag F, Kurth F, Sander LE, et al. SARS-CoV-2 proteome-wide analysis revealed significant epitope signatures in COVID-19 patients. *Front Immunol*. 2021;12:629185.
53. Phan IQ, Subramanian S, Kim D, Murphy M, Pettie D, Carter L, et al. In silico detection of SARS-CoV-2 specific B-cell epitopes and validation in ELISA for serological diagnosis of COVID-19. *Sci Rep*. 2021;11:4290.
54. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and Bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe*. 2020;27:671–80 e2.
55. He W, Shu C, Li B, Zhao J, Cheng Y. Human LINE1 endonuclease domain as a putative target of SARS-associated autoantibodies involved in



- the pathogenesis of severe acute respiratory syndrome. *Chin Med J*. 2008;121:608–14.
56. Freimanis G, Hooley P, Ejtehad HD, Ali HA, Veitch A, Rylance PB, et al. A role for human endogenous retrovirus-K (HML-2) in rheumatoid arthritis: investigating mechanisms of pathogenesis. *Clin Exp Immunol*. 2010;160:340–7.
  57. Talal N, Dauphinée MJ, Dang H, Alexander SS, Hart DJ, Garry RF. Detection of serum antibodies to retroviral proteins in patients with primary sjögren's syndrome (autoimmune exocrinopathy). *Arthritis Rheum*. 1990;33:774–81.
  58. Dang H, Dauphinée MJ, Talal N, Garry RF, Seibold JR, Medsger TA, et al. Serum antibody to retroviral gag proteins in systemic sclerosis. *Arthritis Rheum*. 1991;34:1336–41.
  59. Talal N, Garry RF, Schur PH, Alexander S, Dauphinée MJ, Livas IH, et al. A conserved idiotypic and antibodies to retroviral proteins in systemic lupus erythematosus. *J Clin Invest*. 1990;85:1866–71.
  60. Bengtsson A, Blomberg J, Nived O, Pipkorn R, Toth L, Sturfel G. Selective antibody reactivity with peptides from human endogenous retroviruses and nonviral poly(amino acids) in patients with systemic lupus erythematosus. *Arthritis Rheum*. 1996;39:1654–63.
  61. Hishikawa T, Ogasawara H, Kaneko H, Shirasawa T, Matsuura Y, Sekigawa I, et al. Detection of antibodies to a recombinant gag protein derived from human endogenous retrovirus clone 4-1 in autoimmune diseases. *Viral Immunol*. 1997;10:137–47.
  62. Mustelin T, Ukadike KC. How retroviruses and Retrotransposons in our genome may contribute to autoimmunity in Rheumatological conditions. *Front Immunol*. 2020;11:593891.
  63. Carter V, LaCava J, Taylor MS, Liang SY, Mustelin C, Ukadike KC, et al. High prevalence and disease correlation of autoantibodies against p40 encoded by long interspersed nuclear elements in systemic lupus erythematosus. *Arthritis Rheum*. 2020;72:89–99.
  64. Hung T, Pratt GA, Sundararaman B, Townsend MJ, Chaivorapol C, Bhangale T, et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science*. 2015;350:455–9.
  65. Mavragani CP, Sagalovskiy I, Guo Q, Nezos A, Kapsogeorgou EK, Lu P, et al. Expression of long interspersed nuclear element 1 Retroelements and induction of type I interferon in patients with systemic autoimmune disease. *Arthritis Rheum*. 2016;68:2686–96.
  66. Kaklamanos A, Belogiannis K, Skendros P, Gorgoulis VG, Vlachoyiannopoulos PG, Tzioufas AG. COVID-19 Immunobiology: lessons learned, New Questions Arise. *Front Immunol*. 2021;12:719023.
  67. Moody R, Wilson K, Flanagan KL, Jaworowski A, Plebanski M. Adaptive immunity and the risk of autoreactivity in COVID-19. *Int J Mol Sci*. 2021;22:8965.
  68. Kohm AP, Fuller KG, Miller SD. Mimicking the way to autoimmunity: an evolving theory of sequence and structural homology. *Trends Microbiol*. 2003;11:101–5.
  69. Oldstone MBA. Molecular mimicry and immune-mediated diseases. *FASEB J*. 1998;12:1255–65.
  70. Barnett LA, Fujinami RS. Molecular mimicry: a mechanism for autoimmune injury. *FASEB J*. 1992;6:840–4.
  71. Damian RT. Molecular mimicry in biological adaptation. *Science*. 1965;147:824.
  72. Damian RT. Molecular mimicry revisited. *Parasitol Today*. 1987;3:263–6.
  73. Xiao Y, Rouzine IM, Bianco S, Acevedo A, Goldstein EF, Farkov M, et al. RNA recombination enhances adaptability and is required for virus spread and virulence. *Cell Host Microbe*. 2016;19:493–503.
  74. Lai MMC, Cavanagh D. The molecular biology of coronaviruses. *Adv Virus Res*. 1997;48:1–100.
  75. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol*. 2010;84:3134–46.
  76. Makino S, Keck JG, Stohlman SA, Lai MM. High-frequency RNA recombination of murine coronaviruses. *J Virol*. 1986;57:729–37.
  77. Sallard E, Halloy J, Casane D, Decroly E, van Helden J. Tracing the origins of SARS-CoV-2 in coronavirus phylogenies: a review. *Environ Chem Lett*. 2021;19:769–85.
  78. Wichman HA, Scott L, Howell EK, Martinez AR, Yang L, Baker RJ. Flying around in the genome: characterization of LINE-1 in Chiroptera. *Special Publ Tex Tech Univ Mus*. 2019;71:379–92.
  79. Yan H, Jiao H, Liu Q, Zhang Z, Wang X, Guo M, et al. ACE2 receptor usage reveals variation in susceptibility to SARS-CoV and SARS-CoV-2 infection among bat species. *Nature Ecology & Evolution*. 2021;5:600–8.
  80. Hayward JA, Tachedjian G. Retroviruses of bats: a threat waiting in the wings? *Mbio*. 2021;12:e01941–21.
  81. Allison AC. Contemporary topics in Immunobiology, volume 3. *Contemp Top Immunobiol*. 1974;3:227–42.
  82. Ring GH, Lakkis FG. Breakdown of self-tolerance and the pathogenesis of autoimmunity. *Semin Nephrol*. 1999;19:25–33.
  83. Penzkofer T, Jäger M, Figlerowicz M, Badge R, Mundlos S, Robinson PN, et al. L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Res*. 2017;45:D68–73.
  84. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol*. 2018;14:e1005944.
  85. Ostrovsky A, Hillman-Jackson J, Bouvier D, Clements D, Afgan E, Blankenberg D, et al. Using galaxy to perform large-scale interactive data analyses—an update. *Curr Protoc*. 2021;1:e31.
  86. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res*. 2011;21:487–93.
  87. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605–12.
  88. Yang L, Nilsson-Payant BE, Han Y, Jaffré F, Zhu J, Wang P, et al. Cardiomyocytes recruit monocytes upon SARS-CoV-2 infection by secreting CCL2. *Stem Cell Rep*. 2021. <https://doi.org/10.1016/j.stemcr.2021.07.012>.
  89. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Möller R, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*. 2020;181:1036–45 e9.
  90. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270–3.
  91. Lee J, Arisi I, Puxeddu E, Mramba LK, Amicosante M, Swaisgood CM, et al. Bronchoalveolar lavage (BAL) cells in idiopathic pulmonary fibrosis express a complex pro-inflammatory, pro-repair, angiogenic activation pattern, likely associated with macrophage iron accumulation. *PLoS One*. 2018;13:e0194803.
  92. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
  93. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14:417–9.
  94. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
  95. Ge SX, Son EW, Yao R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. 2018;19:534.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.