Human endogenous retrovirus K activation in the lower respiratory tract of severe COVID-19 patients associates with early mortality

Jairo Temerozo
Oswaldo Cruz Institute - IOC/FIOCRUZ https://orcid.org/0000-0002-8092-2149
Natalia Fintelman-Rodrigues
Fiocruz
Monique Cristina Santos
Fiocruz
Eugênio Hottz
UFJF
Carolina Sacramento
Fiocruz
Aline Silva
Fiocruz
Samuel Mandacaru
Fiocruz
Emilly Caroline Moraes
Fiocruz
Monique Trugilho
Fiocruz
João Gesto
Fiocruz
Marcelo Ferreira
Fiocruz
Felipe Betoni
Fiocruz
Remy Martins-Gonçalves
Fiocruz
Isacláudia Azevedo-Quintanilha
Fiocruz
Juliana Abrantes
UFRJ
Cassia Righy

Fiocruz
Pedro Kurtz
IECPN
Hui Jiang
MGI Tech
Hongdong Tan
MGI Tech
Carlos Morel
Fiocruz
Dumith Bou-Habib
Fiocruz
Fernando Bozza
Oswaldo Cruz Foundation
Patricia Bozza
FIOCRUZ
Thiago Souza (souzatml@gmail.com)
Oswaldo Cruz Foundation

Article

Keywords: COVID-19, SARS-CoV-2, HERV-K

DOI: https://doi.org/10.21203/rs.3.rs-514541/v1

Human endogenous retrovirus K activation in the lower respiratory tract of severe COVID-19 patients associates with early mortality.

Jairo R. Temerozo[a,b,c,d,1], Natalia Fintelman-Rodrigues[c,d,1], Monique Cristina dos Santos[c], Eugenio D. Hottz[c,e], Carolina Q. Sacramento[c,d], Aline de Paula Dias da Silva[c,d], Samuel Coelho Mandacaru[d,f], Emilly Caroline dos Santos Moraes[d,f], Monique R.O. Trugilho[d,f], João S. M. Gesto[d], Marcelo Alves Ferreira[d], Felipe Betoni Saraiva[g], Remy Martins-Gonçalves[c], Isacláudia Gomes Azevedo-Quintanilha[c], Juliana L. Abrantes[h], Cassia Righy[i,j], Pedro Kurtz[i,k], Hui Jiang[l], Hongdong Tan[l], Carlos Morel[d], Dumith Chequer Bou-Habib[a,b], Fernando A. Bozza[j,k], Patrícia T. Bozza[c] and Thiago Moreno L. Souza[c,d]

[a]Laboratory on Thymus Research, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil; [b]National Institute for Science and Technology on Neuroimmunomodulation (INCT/NIM), Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil; [c]Laboratory of Immunopharmacology, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil; [d]National Institute for Science and Technology on Innovation on Neglected Diseases (INCT/IDN), Center for Technological Development in Health (CDTS), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil; [e]Laboratory of Immunothrombosis, Department of Biochemistry, Federal University of Juiz de Fora (UFJF), Minas Gerais, Brazil; [f]Laboratory of Toxinology, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil; [g]Instituto de Tecnologia em Imunobiológicos (Bio-Manguinhos), Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil; [h]Instituto de Ciências Biomédicas, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brazil; [i]Paulo Niemeyer State Brain Institute (IECPN), Rio de Janeiro, RJ, Brazil; [j]Evandro Chagas National Institute of Infectious Diseases, Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, RJ, Brazil; [k]D'Or Institute for Research and Education, Rio de Janeiro, RJ, Brazil; [l]MGI Tech Co. Ltd., Building No.11, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China.

[1]J.R.T. and N.F.-R. and contributed equally to this work.

Abstract

Critically ill 2019 coronavirus disease patients (COVID-19) under invasive mechanical ventilation (IMV) are 10- to 40-times more likely to die than the general population. Although progression from mild to severe COVID-19 has been associated with hypoxia, uncontrolled inflammation and coagulopathy, the mechanisms involved in progression to severity are poorly understood. By analyzing the virome from tracheal aspirates (TA) of 25 COVID-19 patients under IMV, we found higher levels and differential expression of human endogenous retrovirus K (HERV-K) genes compared to nasopharyngeal swabs from mild cases and TA from non-COVID patients. Proteomic analysis and RT-PCR confirmed the presence of HERV-K in these patients. Moreover, increased HERV-K expression was triggered in human primary monocytes from healthy donors after experimental SARS-CoV-2 infection in vitro. In critically ill patients, higher HERV-K levels were associated with early mortality (within 14 days) in the intensive care unit. Increased HERV-K expression in deceased patients associated with IL-17-related inflammation, monocyte activation and higher consumption of clotting/fibrinolysis factors. Our data implicate the levels of HERV-K transcripts in the outcome of critical COVID-19 patients under invasive mechanical ventilation.

Main Text

Severe acute respiratory coronavirus 2 (SARS-CoV-2), the etiological agent of 2019 coronavirus disease (COVID-19), continuously circulates and has caused over one 100,000/month since its original emergence into the human population[1]. Based on official statistics on laboratory-confirmed reports, the case fatality ratio of COVID-19 ranges from 1.5% to 10% in developed and developing countries, respectively[1]. Differently than other highly pathogenic coronaviruses from the 21st century, such as SARS-CoV and Middle East respiratory coronavirus (MERS-CoV), SARS-CoV-2 shedding occurs from the pre-symptomatic period to a few weeks after symptoms onset[2]. Longer viral replication favors tissue damage, as shown by the positive correlation between high activity of lactate dehydrogenase (LDH) activity, a marker of cell death, with COVID-19 progression[3]. While type II pneumocytes are targeted and destroyed by the infection and the respiratory parenchyma is harmed, innate and adaptive immunological responses are not always able to prevent further progression to poor clinical outcomes, and may even worsen the tissue lesions[4,5]. In fact, severe COVID-19 has been associated with increased and uncontrolled release of pro-inflammatory mediators (cytokine storm), so that the resolutive mechanisms are overcome by a marked upregulation of IL-6, TNF-alpha and IL-1-beta[4,5]. In addition, immune cells that orchestrate the innate and adaptive response, such as monocytes and neutrophils, undergo pyroptosis and netosis[6,7]. Consistently, leukopenia and an uncontrolled coagulopathy, marked by platelet activation and high D-dimer levels, correlate with COVID-19 severity[8,9]. Altogether, SARS-CoV-2-triggered inflammation and hypercoagulability have rapidly been defined as main features of the natural history of disease progression from mild to severe COVID-19 clinical presentations[4].

Up to now, the factors described above have been associated with disease progression from mild to severe, but they are limited to explain the mortality of critically ill COVID-19 patients. Further investigation is thus necessary to search for overlooked factors associated with these high mortality rates. Although the stay of COVID-19 patients in the ICU for weeks is more likely to predispose them to nosocomial infections, mortality is high even for patients negative for bacterial infections[10–14]. Despite the best clinical practice to routinely surveillance of bacterial infection in the ICU, unculturable and unbiased diagnosed viruses are neglected in daily practice. The systematic analysis of the virome from critically ill COVID-19 patients is thus necessary, especially in samples from the lower respiratory tract. Thus, we analyzed a cohort of critically ill COVID-19 patients under IMV with sustained SARS-CoV-2 loads, inflammation and coagulopathy to determine whether their lower respiratory tract virome, beyond coronavirus, could improve the rationalization of patients' outcome.

From March to December 2020, we prospectively included 25 critically ill COVID-19 patients requiring IMV, at the median age of 57-year-old and presenting the most common infection symptoms and comorbidities **(Extended Table 1**). Patients displayed high SARS-CoV-2 RNA levels (median of 10^6 copies/mL), laboratory markers of systemic inflammation and coagulopathy (because of elevated plasma levels of C reactive protein [CRP] and D-dimer, respectively), and case fatality ratio of 58% **(Extended Table 1)**. Due to the IMV, the tracheal aspirate (TA) was the source of samples to perform SARS-CoV-2 RNA quantification and virome analysis. We were surprised that the TA of over 70% of the patients had higher SARS-CoV-2 RNA loads than other samples from the lower respiratory tract[15] (**Extended Figure 1A**). RNA content from TAs was unbiased sequenced and rendered an average of $2x10^7$ genomic reads, of which 10% and 90% were virus- or human-related, respectively (**Figure 1A**). Approximately 95% of the virus-associated reads

in the transcriptome were linked to SARS-CoV-2 (**Figure 1A**). Nevertheless, we enriched the new coronavirus sequences using Atoplex kit (MGI, China) **(Supplementary Table 1)** to phylogenetically classify them into the emerging clades 19A, 20A and 20B, in the proportions of the 16%, 12% and 72%, respectively (**Extended Figure 1B and C**), reconfirming that the entire cohort was composed of COVID-19 patients. The SARS-CoV-2 emerging clades identified here were representative of the virus circulating in Brazil during the year of 2020 (https://nextstrain.org/ncov/global?dmax=2020-12-16&dmin=2020-01-16&f_country=Brazil).

In addition to SARS-CoV-2, human endogenous retrovirus K (HERV-K; also known as HML-2) sequences were consistently detected in the TA from these COVID-19 patients, at a proportion of 5 ± 2% (mean ± SEM) of the virome **(Figure 1A and Supplementary Table 2).** Not so important, random phage types were detected in some samples with low coverage **(Figure 1A)**. Among all HERVs, HERV-K is the most contemporaneous in human genome, being incorporated during human-chimpanzee speciation[16]. Thus, it is noteworthy to find an human-specific marker associated with critically ill COVID-19 patients, as non-human primates are less likely to die from SARS-CoV-2 infection[17], raising the attention to a possible role of HERV-K in the dichotomy of SARS-CoV-2 severity between humans and non-human primates.

Moreover, HERV-K was 5-fold more present in the virome of TA aspirates from COVID-19 patients under IMV than in nasopharyngeal swabs (NS) from mild cases, previously studied by us[18] **(Figure 1B and Supplementary Table 2).** HERV-K levels in the TA from deceased or discharged COVID-19 patients did not differ, but these values were significantly higher compared to TA from non-COVID-19 patients, randomly retrieved from sequence read archive (SRA) **(Figure 1B)**. The data from SRA indicate that HERV-K may

be found in the lower respiratory tract of some patients with other illnesses **(Figure 1B**). Indeed, HERV-K detection in the respiratory tract has been associated with lung adenocarcinoma[19], as well as other types of cancer, neurological disorders, multiple sclerosis and arthritis[20]. Here we described that higher levels of HERV-K were found in COVID-19 patients that deceased earlier (**Figure 1C**), accentuating its association with disease severity. Of note, no statistically significant association was found between HERV-K and other features, such as days from COVID-19 onset, age, gender or SARS-CoV-2 RNA levels (**Extended Figure 2**).

Thousands of loci in the human genome are associated with HERV-K-related genetic elements and some of them are actively expressing retroviral structural genes[21], we search for correlation between the HERV-K transcript consensus described here and active HERV-K loci in the human genome. Most often sequences from HERV-K structural genes were expressed from different chromosomic regions, suggesting the activation of otherwise silent genes (**Extended Table 2 and Supplementary Figure 1**). Indeed, critically ill COVID-19 patients differently expressed HERV-K-associated structural genes, doubling and tripling the *gag-pro-pol* and *env* transcripts compared to NS from mild cases and non-COVID TA SRAs **(Figure 1D)**. Expression of *gag,* in prostate cancer and multiple sclerosis has been associated with immune dysregulation[22,23]. The gene *pol* encodes for a reverse transcriptase that may jeopardize the cell cycle of lymphocytes because of its association with leukemia[24]. Gene *pro* is both a protease, which can directly cleave cellular protein with major functional impact in cell biology[25], and a deoxyuridine triphosphate nucleotidohydrolase (dUTPase), whose activity leads to inflammation, immune dysregulation, and progressive obliterative vascular remodeling in the respiratory tract[26]. HERV-K *env* may trigger cell-cell fusion, leading to epithelial to mesenchymal transition and various types of cancer, including in the respiratory tract[19,27].

For further confirmation of HERV-K presence, we performed orthologue assays. TA from 14 patients were enough for further examination via shotgun proteomics **(Figure 2A and B, and Supplementary File 1)**; which revealed more than 20,000 peptides linked to 2,249 human proteins, with a high degree of confidence (FDR <1 at the level of PSM, peptides and proteins identification). To trace similarities with the HERV-K proteome, we compared the peptides from tracheal aspirate human proteome and HERV-K proteins Gag, Pro, Pol, Env and Rec (Uniprot IDs # P62684, P63121, P63132, Q902F9 and P61574, respectively) through BlastP (NCBI/BLAST). A total of 167 alignments were detected with ≥50% coverage and sequence identity, assuming a minimum of 10 residues aligned and peptides ranging in size from 20 to 47 amino acids, under BlastP algorithm default parameters. **(Figure 2 A and B, and Supplementary File 2)**. Besides the proteic detection, *gag* transcripts were quantified by delta-cycle threshold (Ct) (*gag* - ribosomal protein L; RPL, as the housekeeping transcript). The *gag* transcript has been chosen because of its specificity to HERV-K[23]. The levels of the HERV-K in the virome of TA are associated with the lower delta-Ct values (which means gag expression closer to the housekeeping RPL) **(Figure 2C)**. After this validation, we evaluated if HERV-K transcripts were detectable in the plasma from these patients. Indeed, *gag* was detected, with Ct values < 50, being more likely to be detected in deceased patients than discharged individuals or healthy donors (HD) **(Figure 2D)**. These data reconfirm HERV-K in our samples at RNA and peptide levels.

For further evidence of casual relationship between SARS-CoV-2 and the expression of endogenous retrovirus, we experimentally infected either human primary monocytes obtained from healthy donors or Calu-3 cells, a lineage that resembles type II pneumocytes, with SARS-CoV-2 and quantified HERV-K. Upon SARS-CoV-2 infection, HERV-K was up-regulated in the monocytes, but not in Calu-3 cells **(Figure 2E)**, in line with other non-viral

stimuli[28] and with human ontogeny[29]. Monocytes, likewise other immune cells, are important during the natural history of COVID-19, either orchestrating the immune response or succumbing due to pyroptosis and cytokine storm[4,6].

We next examined a possible correlation between HERV-K in the TA with immune-modulation and/or coagulopathy. For this purpose, Spearman correlation analysis for levels of cytokines, coagulation factors and immune cell counts were scored in deceased and discharged patients (**Extended Figure 3**). To be conservative when assuming statistical significance, we additionally performed regression analysis, for those markers that passed Spearman correlation, evaluating differences in angular and/or linear coefficients (**Figure 3**). As a general tendency for the endogenous mediators, HERV-K reduced their levels in the TA **(Extended Figure 3A)** and favored inflammation in the peripheral plasma **(Extended Figure 3B)**. HERV-K negatively associates with two survival/growth factors for immune cells in the blood, granulocyte colony-stimulating factor (G-CSF)[30] and nerve growth factor (NGF)[31] **(Figure 3A)**. Consistently, surviving patients presented higher NGF levels than those who died **(Figure 3A)**. Surprisingly, HERK-K and other HERVs had been reported as inducers of both G-CSF[32] (in immune cells) and NGF[33] (in neurons), however, HERV activation in CNS is associated mainly with the onset and progression of neurological diseases[34], and the induction of G-CSF was evaluated only with a domain of the HERV-K TM protein[32]. As a function of HERV-K levels, other regulatory/anti-inflammatory signals were also decreased in the plasma of deceased, such as IL-1Ra and IL-13 **(Figure 3A)**, which respectively antagonizes IL-1-dependent stimulus and favors an allergenic-like/TH2 response[35,36]. Interesting, the reduction of IL-13 production is reported also by a HERV-H-LTR-derived protein, together with the inhibition of CD4 and CD8 T cell responses[37]. Deceased patients respond to higher HERV-K levels increasing IL-17 **(Figure 3A)**, a further pro-inflammatory mediator that may upregulate IL-6, CRP and airway

remodeling[38], and it is also described to be upregulated by HERV-K and others HERVs in autoimmune diseases[39,40].

During severe COVID-19, clotting factors are intensely consumed[9] and HERV-K levels associated with specific modulations in survivors and deceased patients **(Extended Figure 3C)**. In the light of HERV-K levels, an apparent higher consumption of factor V occurs, being more intense in patients that have deceased **(Figure 3B)**. As clotting cascades are activated, the fibrinolysis product, D-dimer, positively correlates with HERV-K levels in patients that died **(Figure 3B)**, interesting, the modulation of coagulation cascade genes by HERV-K is also described in immune cells[32]. To correlate with cell-mediated immunity specific populations were quantified by flow cytometry **(Supplementary Figure 2)** and plotted as a function of HERV-K levels **(Extended Figure 3D-H)**. HERV-K negatively correlates with natural killer cells **(Figure 3C)**, suggesting a possible contribution to impair an adequate innate antiviral response[41]. Monocyte activation positively correlates with HERV-K **(Figure 3C)**, which agrees with experimental SARS-CoV-2 infection in monocytes, pyroptosis of these cells and release of pro-inflammatory factors[6,42]. These data indicate that HERV-K levels contributed to distinguish discharged from deceased critically ill COVID-19 patients under IMV, reconfirming classical markers of COVID-19 severity and immune activation, and aggregating new information on other potential targets for intervention.

Although HERV-W was shown to correlate with activation and inflammatory markers in blood cells from COVID-19 patients[43], it is not clear whether the expression of HERV-W was due to SARS-CoV-2 or the broad inflammation. We were able to show increased expression of HERV-K due to SARS-CoV-2 infection in monocytes and further correlations with immune activation, coagulopathy and death. HERV-K has been associated with immune activation in HIV-1-infected individuals. For HIV-1-infected antiretroviral therapy-suppressed individuals, augmentation of HERV-K levels precedes immune activation and

HIV-1 rebound[44]. There are different possible links between HERV-K and immune activation, such as Toll-like receptors engagement[45] and direct cell-cell fusion during viral infection[46]. Our data adds SARS-CoV-2 as a trigger of HERV-K expression, with possible correlation with mortality in the ICU. There are also hypotheses linking transposons and retrotransposons to COVID-19 pathogenesis on immune activation and coagulation/fibrinolysis cascade[47,48]; and recently, it was reported the presence of SARS-CoV-2 sequences integrated in host genome, possibly mediated through LINE-1 retrotransposons[49]. Together, these data reinforce our findings, as the pathways for activation of transposable elements are similar, and HERV-K could be concomitantly with these elements, or its activation be the trigger or consequence of SARS-CoV-2 integration.

To the best of our knowledge, our work is the first evidence of the presence of HERV-K in the respiratory tract and in the plasma of critically ill COVID-19 patients. We also connected HERV-K levels with the pro-inflammatory and regulatory events that contribute for patient`s outcome. Our data also gives insights that HERV-K expression is a result of a broader gene expression event during COVID-19, as judged by different activated loci and its expression may further contribute to epigenetic remodeling and long-term consequences. In conclusion, our findings provide original evidence that SARS-CoV-2 triggers increased HERV-K expression, and high levels of HERV-K are associated with disease severity and early mortality.

Methods

Ethics and Patients

From March to December 2020, inpatients from the D'or Institute (ID'or) and Instituto Estadual do Cérebro Paulo Niemayer (IECPN) admitted in the ICU were included upon signed informed consent by their responsible relative. Both TA and Acid-citrate-dextrose (ACD)-anticoagulated blood samples were collected. All patients already had SARS-CoV-2 positive RT-PCR upon entrance in this ward. Nevertheless, we reconfirmed COVID-19 laboratory diagnosis, and summary data from the patients are presented in **Extended Table 1**. The National Review Board of Brazil approved the study protocol (Comissão Nacional de Ética em Pesquisa [CONEP] 30650420.4.1001.0008).

RNA extraction and RT-PCR

The total RNA from TA was extracted using QIAamp Viral RNA (Qiagen, Germany), according to manufacturer's instructions. Quantitative RT-PCR was performed using GoTaq Probe qPCR and RT-qPCR Systems (Promega, USA) in a StepOne Real-Time PCR System (Thermo Fisher Scientific, CA, USA). Primers, probes, and cycling conditions used to detect the SARS-CoV-2 RNA have been described elsewhere[15]. A standard curve was employed for virus quantification, using synthetic RNA for gene N (Microbiologics, MN, USA). Amplifications were carried out in 25 µL reaction mixtures containing 2× reaction mix buffers, 50 µM of each primer, 10 µM of probe, and 5 µL of RNA template.

HERV-K was amplified as described elsewhere[23]. Total RNA from the plasma or culture supernatant was extracted with QIAamp Viral RNA (Qiagen, Germany). Total RNA concentration was determined by spectrophotometric (NanoDrop 2000, ThermoFisher Scientific, CA, USA) and 10 µg of RNA samples were submitted to First-Strand cDNA

synthesis. cDNA was synthesized by 0.5 µl of oligo (dT)20, 0.5 µl of random hexamer primers, 10 mM dNTPs, First-Strand Buffer, 0.1 M DTT and 200U SuperScript III First-Strand Synthesis System (Invitrogen, ThermoFisher Scientific, CA, USA) according to manufacturer's instructions. The total cDNA concentrations were determined by spectrophotometric (NanoDrop 2000, ThermoFisher Scientific, CA, USA). Then, 40-cycle real time PCR was performed using PowerUp SYBR Green Master Mix (Applied Biosystems, ThermoFisher Scientific, CA, USA) in a StepOne Real-Time PCR System (Thermo Fisher Scientific, CA, USA). Amplifications were carried out in 20 µL reaction mixtures containing 10 µL of 2× reaction master mix, 800 nM of each primer (Forward 5'-GCCTTCACATATTCTGTAATT-3' and Reverse 5' -GCAAGGTTGCAAGATGCAGCTC-3') and 100 ng of cDNA following manufacturer's guidelines.

Enrichment-dependent SARS-CoV-2 sequencing

Total viral RNA from TA was extracted and quantified with the QIAamp Viral RNA (Qiagen, Germany) and the Qubit RNA BR Assay Kit (Thermo Fisher Scientific, CA, USA), respectively. cDNA libraries were constructed with the ATOPLex SARS-CoV-2 Full Length Genome Panel v1.0 (kindly donated by MGI Tech Co., Shenzhen, China), an amplicon-based strategy to improve sequencing readout. Dual-indexed, single-stranded library pools were converted to DNA nanoballs by rolling circle amplification and submitted to pair-end sequencing (100nt) on the MGISEQ-2000 platform (recently named as DNBSEQ-G400, MGI Tech Co. Ltd., Shenzhen, China).

Genomic sequences were quality-scored, filtered, trimmed, assembled in contigs, and assigned an ID through a validated workflow for SARS-CoV-2[50] (Genome Detective). Consensus fasta sequences were aligned with ClustalW in Unipro UGENE[51] (version 38),

and phylogenies were constructed with Nextclade[52] to assign the emerging clades **(Supplementary Table 1)**.

Unbiased RNA-Seq

For an unbiased RNA-seq, metatranscriptomics approach, total viral RNA samples were applied to the MGIEasy RNA Library Prep Set (MGI Tech Co. Ltd., Shenzhen, China). Briefly, RNA was initially fragmented by size (250 bp), reverse-transcribed to DNA and added a second strand. Subsequent steps included the end-repair, adaptor-ligation, PCR amplification (for augmenting the overall library yield), denaturation and circularization of single-stranded libraries. Pooled libraries were then converted to DNA nanoballs by rolling circle amplification, and pair-end sequenced (150nt) on the MGISEQ-2000 plataform (MGI Tech Co. Ltd., Shenzhen, China). Virome composition was determined through a validated bioinformatics work-flow[50] (Genome Detective). Raw data were filtered to remove potential adaptors, as well as low-quality and non-viral reads. De novo assembled contigs were compared with reference virus databases (NCBI RefSeq) for obtaining similarity indexes and assigning the species ID. Consensus fasta sequences were generate with the built-in default algorithm (i.e., most frequent base for each alignment position) in Unipro UGENE[51] (version 38) using BAM files generated through Genome Detective[50] work-flow.

HERV-K sequences from polymerase, gag and env were compared with representative genomes deposited in GenBank and three evolutionary analyses were conducted in MEGA X[53] with a total of 100 bootstraps **(Supplementary Figure 1)**. The models for evolutionary analyses were selected upon model-fitting simulation. Models with BIC scores (Bayesian information criterion) were considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (InL), and the number of parameters (including branch lengths)

were obtained. Gag, pol and env evolutionary history was inferred using the maximum-likelihood method and Tamura-Nei model using a discrete Gamma distribution, General Time Reversible model using a discrete Gamma distribution and Hasegawa-Kishino-Yano model using a discrete Gamma distribution, respectively.

Cell, virus and experimental infection

Human lung epithelial cells (Calu-3) and African green monkey kidney cells (Vero E6) were cultured in high glucose DMEM complemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin and 100 µg/mL streptomycin (P/S) at 37 °C in a humidified atmosphere with 5% CO_2. Human primary monocytes were obtained after 3 h of plastic adherence of peripheral blood mononuclear cells (PBMCs). PBMCs were isolated from healthy donors by Ficoll density gradient centrifugation. PBMCs (2 x 10^6 cells) were plated onto 48-well plates in RPMI-1640 without serum for 2 to 4 h. Non-adherent cells were removed, and the remaining monocytes were maintained in DMEM with 5% human serum (HS) and P/S. The purity of human monocytes was above 95%, as determined by flow cytometric analysis (FACScan; Becton Dickinson) using anti-CD3 and anti-CD16 monoclonal antibodies.

SARS-CoV-2 (GenBank # MT710714) was expanded in Vero E6 cells at MOI of 0.01. All procedures related to virus culture were handled in a biosafety level 3 (BSL3) multi-user facility according to WHO guidelines (https://www.who.int/publications/i/item/WHO-WPE-GIH-2021.1). Virus titers were determined as plaque forming units (PFU)/mL. Virus stocks were kept in -80 °C ultra-low freezers.

Infections were performed with SARS-CoV-2 at MOI of 0.01 (monocytes) or 0.1 (Calu-3) in low (monocytes) or high (Calu-3) glucose DMEM without serum. After 1 hour, cells were washed and incubated with complete medium. After 24h (monocytes) or 48h (Calu-3), culture supernatant was harvested for HERV-K quantification.

Flow Cytometry

Whole blood samples were incubated for 10 min with FACS lysing buffer (BD Biosciences) and then centrifuged at 400 RCF for 15 min. The supernatant was discarded and cells were resuspended in HEPES-Tyrode (HT) buffer (10 mM HEPES, 137 mM NaCl, 2.8 mM KCl, 1 mM $MgCl_2$.6H2O, 12 mM $NaHCO_3$, 0.4 mM Na_2HPO_4, 5.5 mM glucose, 0.35% BSA [pH 7.4]). Monocytes were labeled with fluorescein isothiocyanate (FITC)-conjugated anti-CD16, phycoerythrin (PE)-conjugated anti-TF and peridinin-chlorophyll (PerCP)-conjugated anti-CD14 (BD Pharmingen), or with FITC-conjugated anti-CD38, PE-conjugated anti-CD11b and PerCP-conjugated anti-CD14, or with FITC-conjugated anti-HLA-DR, PerCP-conjugated anti-CD14 and allophycocyanin (APC)-conjugated anti-CD83 (BD Pharmingen). Lymphocytes were labeled with FITC-conjugated anti-CD3, PE-conjugated anti-CD4 and APC-H7-conjugated anti-CD8, or with FITC-conjugated anti-CD11b, PE-conjugated anti-CD25, PE-Cy5-conjugated anti-CD38 and APC-H7-conjugated anti-CD8. B cells were labeled with FITC-labeled anti-CD38, PE-conjugated anti-CD19, PerCP-conjugated anti-CD20 and APC-H7-conjugated anti-CD27. NK cells were labeled with FITC-conjugated anti-CD107, PE-conjugated anti-CD11b, PE-Cy5-conjugated anti-CD56, APC-conjugated anti-CD3 and APC-H7-conjugated anti-CD27. Neutrophils were labeled with FITC-conjugated anti-myeloperoxidase (MPO) and PE-conjugated anti-CD11b. Cells were incubated with antibodies for 30 min at room temperature and fixed with 4% paraformaldehyde. Cells labeled with each antibody separately were used for appropriate

color compensation and isotype-matched IgG conjugated with the same fluorochromes were used as the negative controls. Lymphocytes, monocytes and neutrophils were recognized by their characteristic forward and side scatter and expression of specific surface markers as shown in **Supplemental Figure S2**. Flow cytometer (BD FACSCalibur) was used to acquire 2,000 to 5,000 gated events. Acquired data were further analyzed using FlowJo software.

Elisa

Blood samples were collected in ACD-containing syringes and plasma was obtained by serial centrifugation. Whole-blood samples were centrifuged at 150 RCF/20min/25 ºC, to obtain platelet-rich plasma (PRP), then 500 RCF/20min/25 °C to obtain platelet-poor plasma (PPP) and finally 2,500 RCF/20min/25 °C to obtain platelet-free plasma, which was then aliquoted into 1 mL samples and conditioned at -80 °C. Commercial ELISA (R&D Systems, MN, USA) and Multiplex (BioRad, CA, EUA) kits were used to measure cytokines, chemokine and coagulation markers.

Proteomic Sample preparation

50 µg of 14 tracheal aspirated samples were lysed in 8M urea solubilized in 20 mM of ammonium bicarbonate pH 7.9 containing a complete mixture of protease and phosphatase inhibitors (Roche, Switzerland). After centrifugation at 14,000 RCF for 20 min, the supernatants were transferred to new tubes and heated at 32 °C for 30 min under 600 rpm agitation. Proteins were reduced with 5 mM of dithiothreitol for 60 min at 32 °C and alkylated in 14 mM of iodoacetamide for 40 min at room temperature in the dark. Samples were then diluted to a 1M urea and 1 µg of modified trypsin (Promega, WI, EUA) (1:50 w/w - trypsin:substrate ratio) was added. Each sample was then incubated for 18 h at 37 °C.

Tryptic peptide were acidified with TFA (0.1% (v/v) final concentration) and desalted with POROS R2 resin (Applied Biosystems, CA, EUA), packaged in micropipette tips (Millipore, Bedford, USA). Desalted peptides were dried, suspended in 10 µl of 0.1% formic acid and aliquots corresponding to 0.5 µg/µl were separated for mass spectrometry analysis.

Mass spectrometry

The tryptic digests were analyzed by reversed-phase nanochromatography coupled to high-resolution nanoelectrospray ionization mass spectrometry. Chromatography was performed using a Dionex Ultimate 3000 RSLCnano system coupled to the HF-X Orbitrap mass spectrometer (Thermo Fischer Scientific, CA, EUA). Samples (1 µg per run) were initially applied to a 2 cm guard column, followed by fractionation on a 25.5 cm PicoFritTM Self-Pack column (New Objective) packed with 1.9 µm silica, ReproSil-684 Pur 120 Å C18-AQ (Dr.Maisch, Germany). Samples were loaded in 0.1% (v/v) formic acid (FA) and 2% acetonitrile (ACN) onto the trap column at 2 µL/min, while chromatographic separation occurred at 200 nL/min. Mobile phase A consisted of 0.1% (v/v) FA in water, while mobile phase B consisted of 0.1% (v/v) FA in ACN. Peptides were eluted with a linear gradient from 2 to 40% eluent B over 32 min, followed by up to 80% B in 4 min. Lens voltage was set to 60 V. Full scan MS mode was acquired with a resolution of 60,000 (FWHM at m/z 200 and AGC set to $3x10^6$). Up to 20 most abundant precursor ions from each scan (m/z 350-1,400) were sequentially subjected to fragmentation by HCD. Fragment ions were analyzed at a resolution of 15,000 using an AGC set to $1x10^5$. Data were acquired using Xcalibur software (version 4.2.47).

Proteomic computational analysis

The raw data files were processed and quantified using PatternLab for Proteomics software[54] (version 4.0). Peptide sequence matching (PSM) was performed using the Comet algorithm against the protein-centric human database neXtProt[55] plus the SARS-CoV-2 reference proteome from Uniprot[56], under ID UP000464024, both downloaded March 29, 2021. A target-decoy strategy was employed. The search parameters were; tryptic and semi-tryptic peptides, with masses between 500 and 5,000 Da, up to 2 lost cleavage sites; modifications: carbamidomethylation (Cys), oxidation (Met) and initial tolerance of 40 ppm for precursor ions. PSMs were filtered using the Search Engine Processor (SEPro) module and identifications were grouped by the number of enzymatically cleaved ends, resulting in two distinct subgroups. For each result, the scores for each metric (XCorr, DeltaCN, and ZScore) were used to generate a Bayesian discriminator, accepting up to 1% false discovery rate (FDR), estimated by the number of decoy sequence IDs. Results were further filtered to accept only PSMs with mass error less than 5 ppm and protein identifications supported by two or more independent identifications. Proteins identified by a single spectrum (1 hit wonder) with XCorr below 2 were excluded. The final list of identified peptides and mapped proteins for all samples was reported. The list of resulting peptides from shotgun proteomics was used for alignment with the sequences of Human endogenous retrovirus K113 (https://www.ncbi.nlm.nih.gov/nuccore/NC_022518.1). The alignment was carried out with the NCBI/BLAST database, through the Protein Blast - BlastP algorithm. Alignments with identity and coverage equal to or greater than 50% were considered. Detailed information about the proteins that aligned with the peptides can be obtained from the UniProtKB SwissProt database

(https://www.uniprot.org/uniprot/?query=Human+endogenous+retrovirus+K113&sort=score)

.

Statistics

The assays were performed blinded by one professional, codified and then read by another professional. All experiments were carried out at least three independent times, including a minimum of two technical replicates in each assay. Prism GraphPad software 8.0 was preferentially used to generate the data sets. One-way analysis of variance (ANOVA) was used to compare differences among 3 or more groups following a normal (parametric) distribution, and Tukey's post-hoc test was used to locate the differences between the groups; or Friedman's test (for non-parametric data) with Dunn's post-hoc test. Spearman correlation was used for comparison of curves, as well as angular and linear comparisons between discharged and deceased patients. Statistically significant p values < 0.05 were registered.

References

1. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).

2. Benefield, A. E. *et al.* SARS-CoV-2 viral load peaks prior to symptom onset: A systematic review and individual-pooled analysis of coronavirus viral load from 66 studies. *medRxiv* 2020.09.28.20202028 (2020) doi:10.1101/2020.09.28.20202028.

3. Henry, B. M. *et al.* Lactate dehydrogenase levels predict coronavirus disease 2019 (COVID-19) severity and mortality: A pooled analysis. *Am. J. Emerg. Med.* **38**, 1722–1726 (2020).

4. Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* **20**, 363–374 (2020).

5. Laing, A. G. *et al.* A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat. Med.* **26**, 1623–1635 (2020).

6. Ferreira, A. C. *et al.* SARS-CoV-2 engages inflammasome and pyroptosis in human primary monocytes. *Cell Death Discov.* **7**, 43 (2021).

7. Veras, F. P. *et al.* SARS-CoV-2-triggered neutrophil extracellular traps mediate COVID-19 pathology. *J. Exp. Med.* **217**, (2020).

8. Terpos, E. *et al.* Hematological findings and complications of COVID-19. *Am. J. Hematol.* **95**, 834–847 (2020).

9. Gómez-Mesa, J. E., Galindo-Coral, S., Montes, M. C. & Muñoz Martin, A. J. Thrombosis and Coagulopathy in COVID-19. *Curr. Probl. Cardiol.* **46**, 100742 (2021).

10. Lim, Z. J. *et al.* Case Fatality Rates for Patients with COVID-19 Requiring Invasive Mechanical Ventilation. *Am. J. Respir. Crit. Care Med.* **203**, 54–66 (2021).

11. Chang, R., Elhusseiny, K. M., Yeh, Y.-C. & Sun, W.-Z. COVID-19 ICU and mechanical ventilation patient characteristics and outcomes—A systematic review and meta-analysis. *PLoS One* **16**, e0246318 (2021).

12. Vanhomwegen, C. *et al.* Procalcitonin accurately predicts mortality but not bacterial infection in COVID-19 patients admitted to intensive care unit. *Ir. J. Med. Sci.* 1–4 (2021) doi:10.1007/s11845-020-02485-z.

13. Soriano, M. C. *et al.* Low incidence of co-infection, but high incidence of ICU-acquired

infections in critically ill patients with COVID-19. *J. Infect.* **82**, e20–e21 (2021).

14. Bardi, T. *et al.* Nosocomial infections associated to COVID-19 in the intensive care unit: clinical characteristics and outcome. *Eur. J. Clin. Microbiol. Infect. Dis.* **40**, 495–502 (2021).

15. Wölfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020).

16. Subramanian, R. P., Wildschutte, J. H., Russo, C. & Coffin, J. M. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* **8**, 90 (2011).

17. Singh, D. K. *et al.* Responses to acute infection with SARS-CoV-2 in the lungs of rhesus macaques, baboons and marmosets. *Nat. Microbiol.* **6**, 73–86 (2021).

18. Fintelman-Rodrigues, N. *et al.* Viral Genetic Evidence and Host Immune Response of a Small Cluster of Individuals with Two Episodes of SARS-CoV-2 Infection. *SSRN Electron. J.* (2020) doi:10.2139/ssrn.3750109.

19. Zare, M. *et al.* Human endogenous retrovirus env genes: Potential blood biomarkers in lung cancer. *Microb. Pathog.* **115**, 189–193 (2018).

20. Xue, B., Sechi, L. A. & Kelvin, D. J. Human Endogenous Retrovirus K (HML-2) in Health and Disease. *Front. Microbiol.* **11**, 1690 (2020).

21. Xue, B. *et al.* Identification of the distribution of human endogenous retroviruses K (HML-2) by PCR-based target enrichment sequencing. *Retrovirology* **17**, (2020).

22. Reis, B. S. *et al.* Prostate cancer progression correlates with increased humoral

immune response to a human endogenous retrovirus GAG protein. *Clin. Cancer Res.* **19**, 6112–6125 (2013).

23. PJ, B. Analysis of Human Endogenous Retrovirus Expression in Multiple Sclerosis Plaques. *J. Emerg. Dis. Virol.* **3**, (2017).

24. Bergallo, M. *et al.* Expression of the pol gene of human endogenous retroviruses HERV-K and -W in leukemia patients. *Arch. Virol.* **162**, 3639–3644 (2017).

25. Rigogliuso, G. *et al.* A human endogenous retrovirus encoded protease potentially cleaves numerous cellular proteins. *Mob. DNA* **10**, 36 (2019).

26. Saito, T. *et al.* Upregulation of Human Endogenous Retrovirus-K Is Linked to Immunity and Inflammation in Pulmonary Arterial Hypertension. *Circulation* **136**, 1920–1935 (2017).

27. Lemaître, C., Tsang, J., Bireau, C., Heidmann, T. & Dewannieux, M. A human endogenous retrovirus-derived gene that can contribute to oncogenesis by activating the ERK pathway and inducing migration and invasion. *PLOS Pathog.* **13**, e1006451 (2017).

28. Johnston, J. B. *et al.* Monocyte activation and differentiation augment human endogenous retrovirus expression: Implications for inflammatory brain diseases. *Ann. Neurol.* **50**, 434–442 (2001).

29. Bergallo, M. *et al.* Human Endogenous Retroviruses Are Preferentially Expressed in Mononuclear Cells From Cord Blood Than From Maternal Blood and in the Fetal Part of Placenta. *Front. Pediatr.* **8**, 244 (2020).

30. Basu, S., Hodgson, G., Katz, M. & Dunn, A. R. Evaluation of role of G-CSF in the production, survival, and release of neutrophils from bone marrow into circulation. *Blood* **100**, 854–861 (2002).

31. Minnone, G., De Benedetti, F. & Bracci-Laudiero, L. NGF and Its Receptors in the Regulation of Inflammatory Response. *Int J Mol Sci* **18**, (2017).

32. Morozov, V. A., Dao Thi, V. L. & Denner, J. The Transmembrane Protein of the Human Endogenous Retrovirus - K (HERV-K) Modulates Cytokine Release and Gene Expression. *PLoS One* **8**, e70399 (2013).

33. Bhat, R. K. *et al.* Human Endogenous Retrovirus-K(II) Envelope Induction Protects Neurons during HIV/AIDS. *PLoS One* **9**, e97984 (2014).

34. Gruchot, J., Kremer, D. & Küry, P. Neural Cell Responses Upon Exposure to Human Endogenous Retroviruses. *Front. Genet.* **10**, 655 (2019).

35. Gabay, C., Lamacchia, C. & Palmer, G. IL-1 pathways in inflammation and human diseases. *Nat. Rev. Rheumatol.* **6**, 232–241 (2010).

36. Zhu, J. T helper 2 (Th2) cell differentiation, type 2 innate lymphoid cell (ILC2) development and regulation of interleukin-4 (IL-4) and IL-13 production. *Cytokine* **75**, 14–24 (2015).

37. Zhao, R. *et al.* HHLA2 is a member of the B7 family and inhibits human CD4 and CD8 T-cell function. *Proc. Natl. Acad. Sci.* **110**, 9879–9884 (2013).

38. Amatya, N., Garg, A. V. & Gaffen, S. L. IL-17 Signaling: The Yin and the Yang. *Trends Immunol.* **38**, 310–322 (2017).

39. Ariza, M.-E. & Williams, M. V. A Human Endogenous Retrovirus K dUTPase Triggers a TH1, TH17 Cytokine Response: Does It Have a Role in Psoriasis? *J. Invest. Dermatol.* **131**, 2419–2427 (2011).

40. Wang, X. *et al.* Increased HERV-E clone 4–1 expression contributes to DNA hypomethylation and IL-17 release from CD4+ T cells via miR-302d/MBD2 in systemic lupus erythematosus. *Cell Commun. Signal.* **17**, 94 (2019).

41. Devulder, J. *et al.* Aberrant anti-viral response of natural killer cells in severe asthma. *Eur. Respir. J.* **55**, (2020).

42. Dias, S. da S. G. *et al.* Lipid droplets fuel SARS-CoV-2 replication and production of inflammatory mediators. *PLOS Pathog.* **16**, e1009127 (2020).

43. Balestrieri, E. *et al.* Evidence of the pathogenic HERV-W envelope expression in T lymphocytes in association with the respiratory outcome of COVID-19 patients. *EBioMedicine* **66**, 103341 (2021).

44. Contreras-Galindo, R., Almodóvar-Camacho, S., González-Ramírez, S., Lorenzo, E. & Yamamura, Y. Short Communication: Comparative Longitudinal Studies of HERV-K and HIV-1 RNA Titers in HIV-1-Infected Patients Receiving Successful versus Unsuccessful Highly Active Antiretroviral Therapy. *AIDS Res. Hum. Retroviruses* **23**, 1083–1086 (2007).

45. Dembny, P. *et al.* Human endogenous retrovirus HERV-K(HML-2) RNA causes neurodegeneration through Toll-like receptors. *JCI Insight* **5**, (2020).

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. **3**, 246 (2013).

47. Li, M., Schifanella, L. & Larsen, P. A. Alu retrotransposons and COVID-19 susceptibility and morbidity. *Hum. Genomics* **15**, 2 (2021).

48. Yin, Y., Liu, X., He, X. & Zhou, L. Exogenous Coronavirus Interacts With Endogenous Retrotransposon in Human Cells. *Front. Cell. Infect. Microbiol.* **11**, (2021).

49. Zhang, L. *et al.* Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proc. Natl. Acad. Sci.* **118**, e2105968118 (2021).

50. Vilsker, M. *et al.* Genome Detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **35**, 871–873 (2019).

51. Okonechnikov, K., Golosova, O. & Fursov, M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).

52. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

53. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

54. Carvalho, P. C. *et al.* Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0. *Nat. Protoc.* **11**, 102–117 (2016).

55. Zahn-Zabal, M. *et al.* The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Res.* **48**, D328–D334 (2019).

56. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

Acknowledgements

We thank Carmen Gripp for assessments related to the BSL3 facility and Marco Alberto Medeiros related to the sequencing platform, and MGI, a partner in the implementation of next-generation sequencing. We thank the Hemotherapy Service from Hospital Clementino Fraga Filho (Federal University of Rio de Janeiro, Brazil) for providing buffy-coats.

Funding

This work was financially supported by the Brazilian agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Inova-Fiocruz Project (B3) and Mercosur Fund for Structural Convergence (FOCEM, Mercosur, grant number 03/11) granted for Thiago Moreno L. Souza, Patrícia T. Bozza, Fernando A. Bozza and Dumith Chequer Bou-Habib. This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) with finance code 001. Funding was also provided by CNPq, CAPES, and FAPERJ through the National Institutes of Science and Technology Program (INCT) to Carlos Morel (INCT/IDPN) and to Wilson Savino (INCT/NIM). The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data, as well as in the writing of the letter, and in the decision to publish the results.

Authors Contributions

C.Q.S., N.F.-R., and J.R.T. performed the cell assays; N.F.-R., A.P.D.S., M.C.S., F.B.S., M.A.F., J.S.M.G., H.J., and H.T. performed sequencing; N.F.-R., J.R.T., M.C.S., A.P.D.S. and J.S.M.G. were responsible for bioinformatics; E.H., R.M.-G., and I.G.A.-Q. performed immunological assessments; S.C.M., E.C.S.M. and M.R.O.T. conducted proteomic assays, C.Q.S., N.F.-R., J.R.T., E.H., J.L.A. and T.M.L.S. analyzed data; F.A.B. and P.T.B. conducted clinical surveillance; F.A.B., C.R. and P.K. coordinated patients enrolment; P.T.B., D.C.B., C.M. and T.M.L.S. handled study coordination; D.C.H., J.R.T., C.M. and T.M.L.S. manuscript preparation; All authors revised and approved the manuscript in final version.

Competing interest declaration: The authors declare no competing interest.

Additional information

Corresponding author line: Thiago Moreno Lopes e Souza, Fundação Oswaldo Cruz (Fiocruz), Centro de Desenvolvimento Tecnológico em Saúde (CDTS), Instituto Oswaldo Cruz (IOC), Pavilhão Osório de Almeida, sala 16, Av. Brasil 4365, Manguinhos, Rio de Janeiro, RJ, Brasil, CEP 21060340; tmoreno@cdts.fiocruz.br

Figure 1

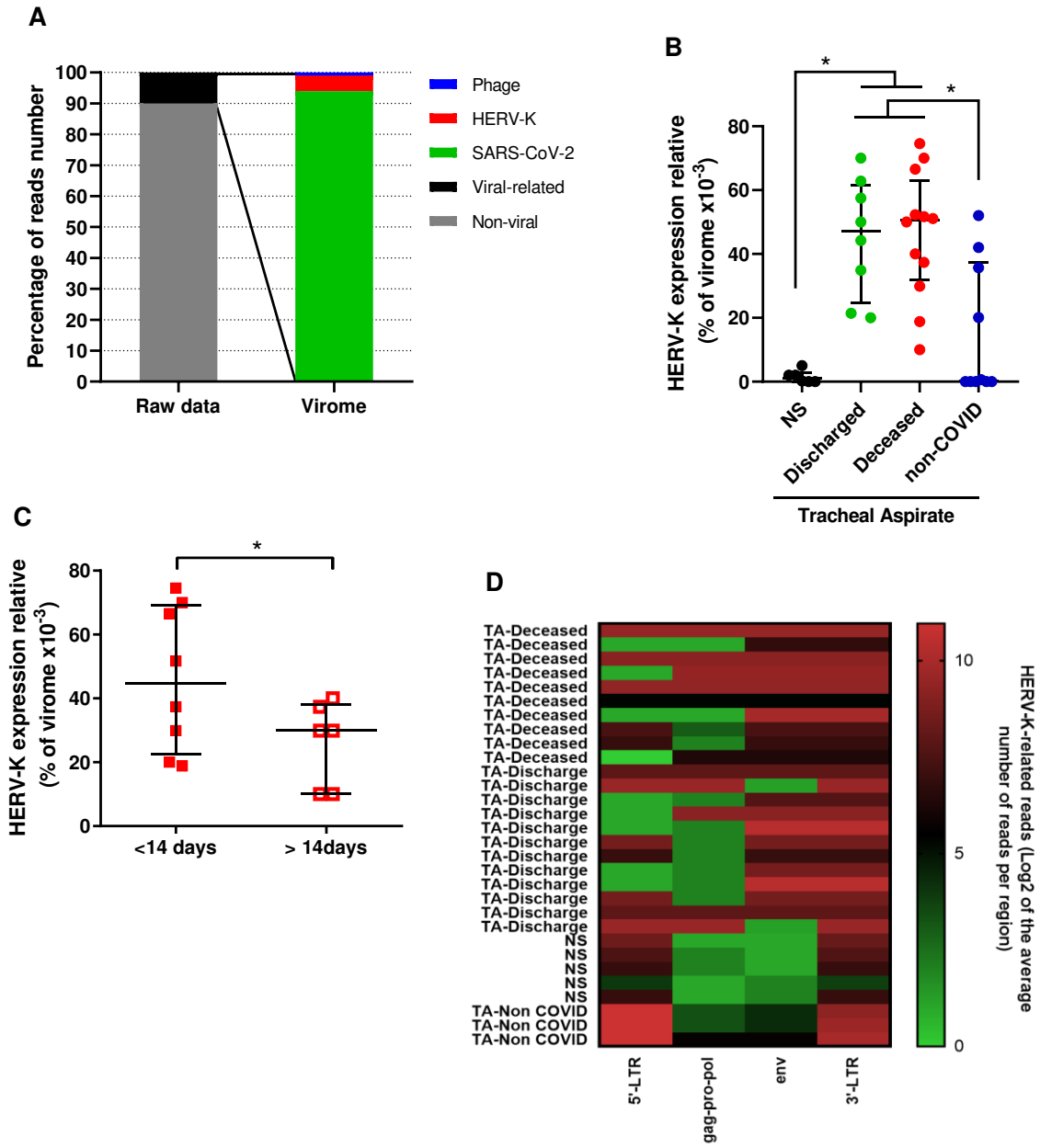

Figure 1 - Differential overexpression of HERV-K transcripts in the lower respiratory tract of critically ill COVID-19 patients associates with early mortality. RNA-sequencing of tracheal aspirates (TA) from severe cases **(Table S1)** and of nasopharyngeal swabs (NS) from mild cases[17] were performed on the MGI-2000 RNAseq platform and high-quality sequences (Q ≥30) were selected for downstream analysis. **A)** Percentage of virus-related reads in the transcriptome and virome from the TA. **B)** The percentage of HERV-K-related reads in the virome, compared to NS and non-COVID TA (# SRX3934910, SRX3934905, SRX4213551, SRX4213553, SRX4213553, SRX4213548, SRX4213540, SRX3934932, SRX3934906 and SRX4213544) **C)** HERV-K expression in TA over time (days) from ICU admission to death. **D)** Absolute HERV-K read counts for TA from critical COVID-19 patients, NS from mild cases and for the three non-COVID TA Sequence Read Archive (SRA) with more pronounced HERV-K signals. * $p<0.05$.

Figure 2

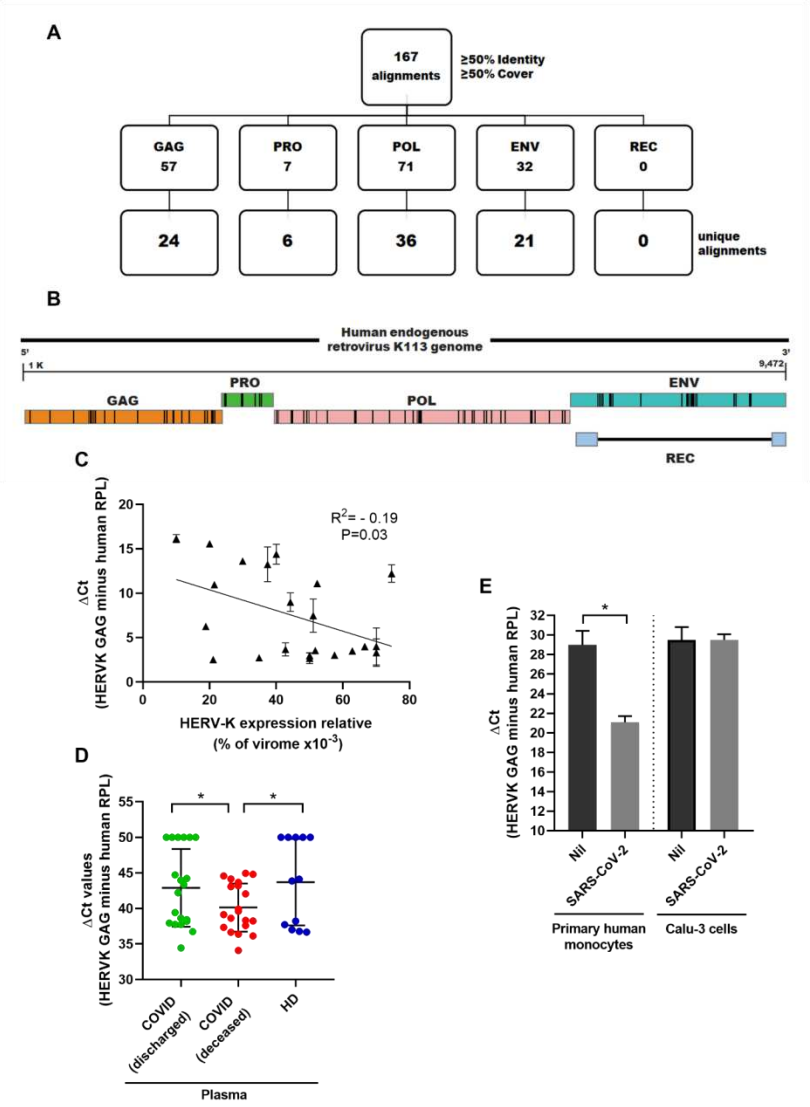

Figure 2 - Orthologue assays to reconfirm COVID-19/SARS-CoV-2 and HERV-K GAG engagement. **A)** Schematic representation of 130 *BlastP* alignments detected between peptides identified in tracheal aspirate human proteome and HERVK proteins. The 130 found alignments are divided as shown: 53 alignments in GAG, 2 in PRO, 54 in POL, 20 in ENV and 1 in REC. Considering the alignments detected exclusively in each HERV-K protein: 29, 2, 37, 15 and none are observed, respectively. **B)** Each alignment is represented as a black bar on HERV-K proteins. **C)** The fraction of the HERV-K virome was compared to the results of real time RT-PCR to detect HERV-K GAG (minus the housekeeping RPL, as the delta-Ct values). **D)** Plasma samples from severe cases (Table S1) and from healthy donors (HD) were evaluated for presence of HERV-K GAG by real time RT-PCR. Samples with Ct values below 40 were considered positive for HERV-K. Delta-Ct was calculated by subtracting RPL from HERV-K GAG Ct values. **E)** Human primary monocytes (n=5, 2 technical replicates) or Calu-3 cells (n=3, 2 technical replicates) were infected with MOI of 0.1. At 24h post-infection, the culture supernatant was harvested and total RNA used to quantify HERV-K GAG and RPL. *p<0.05.

Figure 3

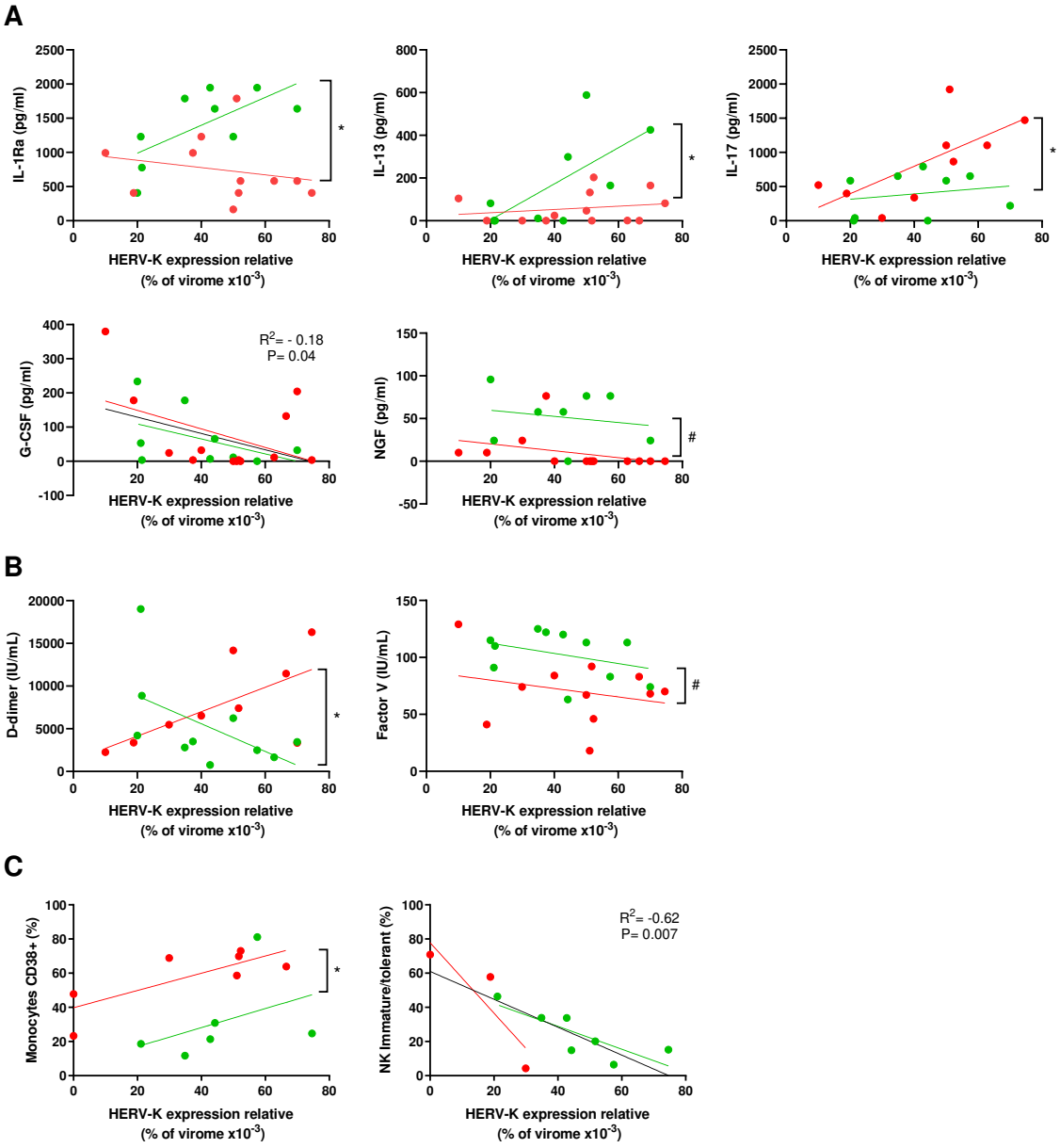

Figure 3 - HERV-K levels correlate with immune activation and coagulopathy in a patient outcome-dependent manner. HERV-K levels are presented as the function of **A)** cellular survival/differentiation factors or interleukins, **B)** clotting or fibrinolysis cascade markers, and **C)** immune cells. These are the statistically significant analyses from **Extended Figure 3** (Panels in **A and B** derived from **Ext. Fig. 3A and 3C**, respectively; panels in **C** derived from **Ext. Fig. 3D and 3H**). Patients and regression lines are highlighted in green ● for discharged and in red ● for deceased patients. Regression lines in black indicate that statistical significance when combining both discharged and deceased patients. Statistically significant ($p<0.05$) differences between linear or angular coefficients are represented by # or *, respectively.

Figures

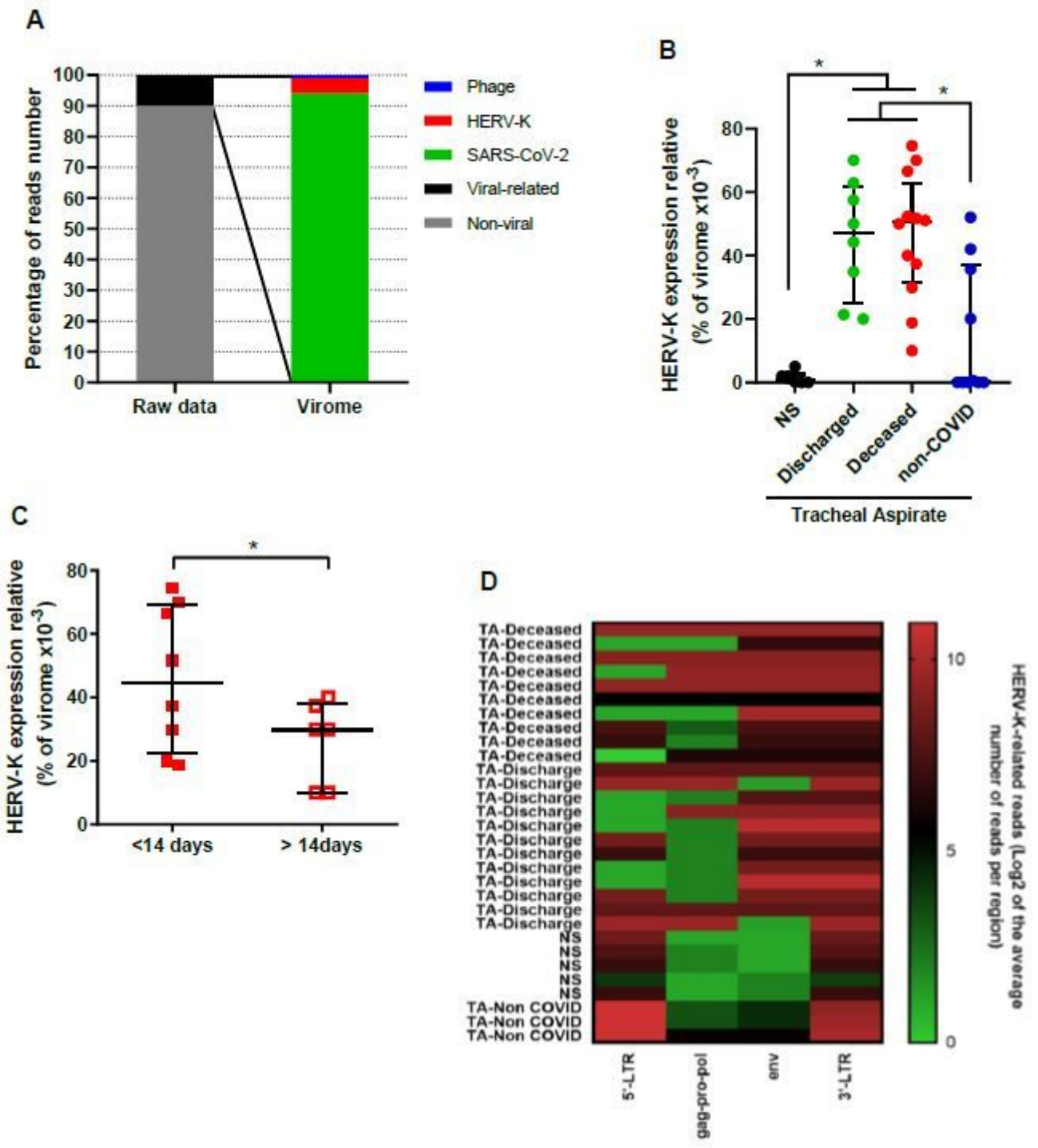

Figure 1

Differential overexpression of HERV-K transcripts in the lower respiratory tract of critically ill COVID-19 patients associates with early mortality. RNA-sequencing of tracheal aspirates (TA) from severe cases (Table S1) and of nasopharyngeal swabs (NS) from mild cases17 were performed on the MGI- 2000 RNAseq platform and high-quality sequences (Q ≥30) were selected for downstream analysis. A) Percentage of virus-related reads in the transcriptome and virome from the TA. B) The percentage of

HERVK- related reads in the virome, compared to NS and non-COVID TA (# SRX3934910, SRX3934905, SRX4213551, SRX4213553, SRX4213553, SRX4213548, SRX4213540, SRX3934932, SRX3934906 and SRX4213544) C) HERV-K expression in TA over time (days) from ICU admission to death. D) Absolute HERV-K read counts for TA from critical COVID-19 patients, NS from mild cases and for the three non-COVID TA Sequence Read Archive (SRA) with more pronounced HERV-K signals. * p<0.05.

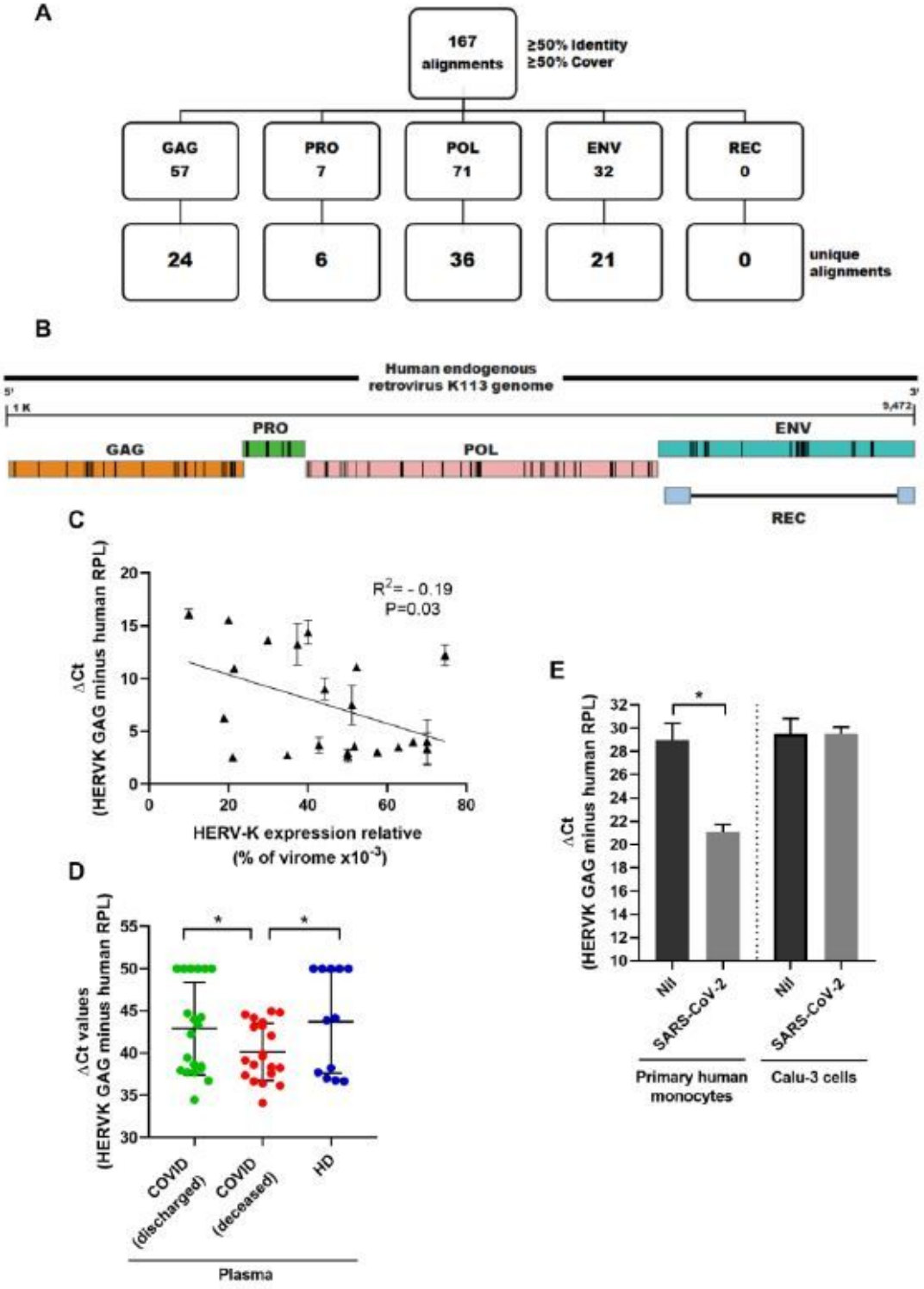

Figure 2

Orthologue assays to reconfirm COVID-19/SARS-CoV-2 and HERV-K GAG engagement. A) Schematic representation of 130 BlastP alignments detected between peptides identified in tracheal aspirate human proteome and HERVK proteins. The 130 found alignments are divided as shown: 53 alignments in GAG, 2 in PRO, 54 in POL, 20 in ENV and 1 in REC. Considering the alignments detected exclusively in each HERV-K protein: 29, 2, 37, 15 and none are observed, respectively. B) Each alignment is represented as a black bar on HERV-K proteins. C) The fraction of the HERV-K virome was compared to the results of real time RT-PCR to detect HERV-K GAG (minus the housekeeping RPL, as the delta-Ct values). D) Plasma samples from severe cases (Table S1) and from healthy donors (HD) were evaluated for presence of HERV-K GAG by real time RT-PCR. Samples with Ct values below 40 were considered positive for HERV-K. Delta-Ct was calculated by subtracting RPL from HERV-K GAG Ct values. E) Human primary monocytes (n=5, 2 technical replicates) or Calu-3 cells (n=3, 2 technical replicates) were infected with MOI of 0.1. At 24h post-infection, the culture supernatant was harvested and total RNA used to quantify HERV-K GAG and RPL. *$p<0.05$.

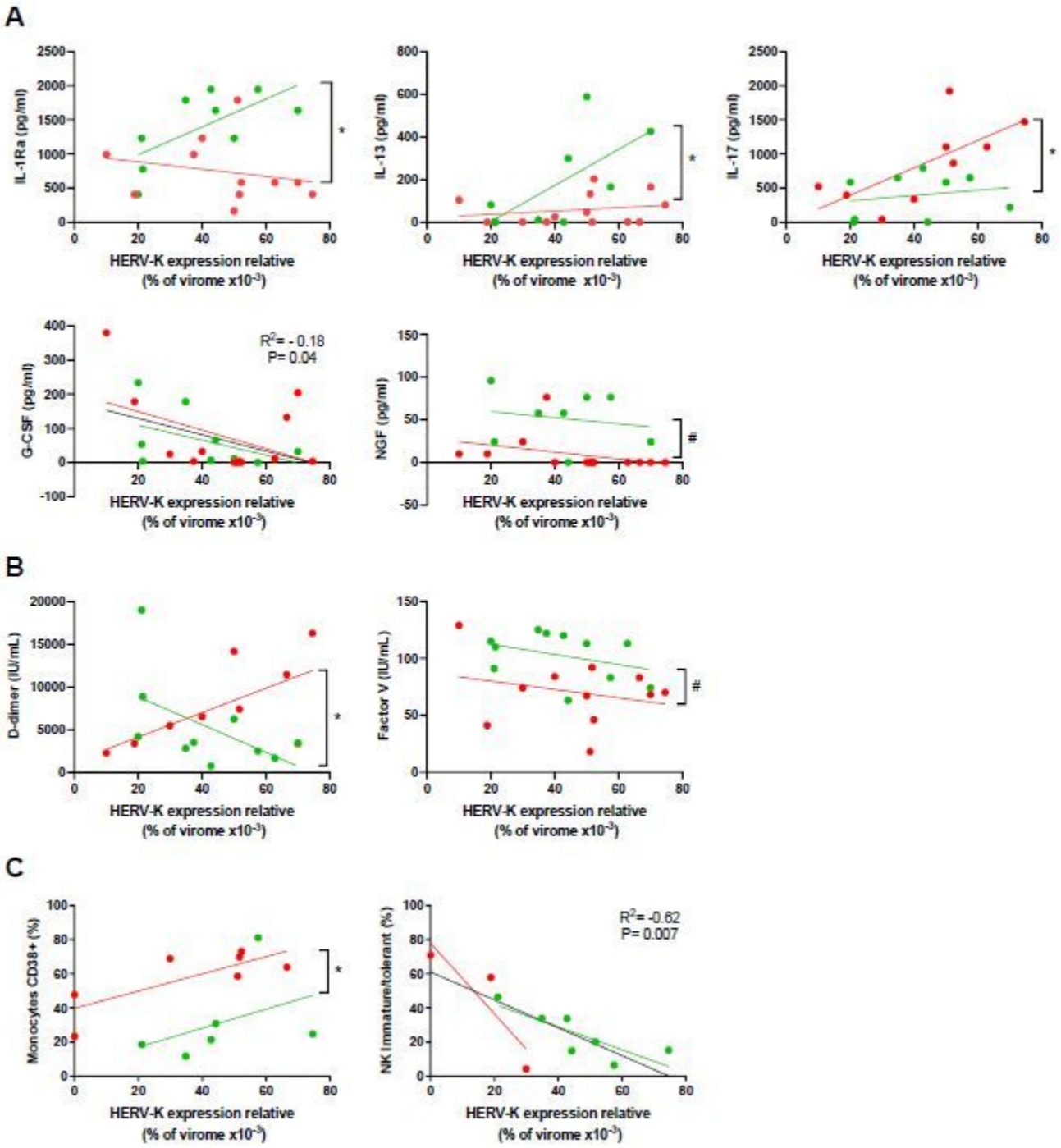

Figure 3

HERV-K levels correlate with immune activation and coagulopathy in a patient outcome-dependent manner. HERV-K levels are presented as the function of A) cellular survival/differentiation factors or interleukins, B) clotting or fibrinolysis cascade markers, and C) immune cells. These are the statistically significant analyses from Extended Figure 3 (Panels in A and B derived from Ext. Fig. 3A and 3C, respectively; panels in C derived from Ext. Fig. 3D and 3H). Patients and regression lines are highlighted in green for discharged and in red for deceased patients. Regression lines in black indicate that statistical

significance when combining both discharged and deceased patients. Statistically significant ($p<0.05$) differences between linear or angular coefficients are represented by # or *, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MSSupplementaryFile1.xlsx
- MSSupplementaryFile2.xlsx
- MSExtendedandSupplementaryData.pdf
- MSExtendedandSupplementaryData.pdf