

# 1 Upregulation of Human Endogenous Retroviruses in Bronchoalveolar Lavage

## 2 Fluid of COVID-19 Patients

3

4 Konstantina Kitsou<sup>1,2</sup>, Anastasia Kotanidou<sup>3</sup>, Dimitrios Paraskevis<sup>1</sup>, Timokratis  
5 Karamitros<sup>4</sup>, Aris Katzourakis<sup>5</sup>, Richard Tedder<sup>6</sup>, Tara Hurst<sup>7</sup>, Spyros Sapounas<sup>8</sup>,  
6 Athanassios Kotsinas<sup>9</sup>, Vassilis Gorgoulis<sup>9,10,11,12</sup>, Vana Spoulou<sup>2</sup>, Sotirios  
7 Tsiodras<sup>13</sup>, Pagona Lagiou<sup>1</sup>, Gkikas Magiorkinis<sup>1,\*</sup>

8

- 9 1. Department of Hygiene, Epidemiology and Medical Statistics, Medical School,  
10 National and Kapodistrian University of Athens, Greece
- 11 2. Immunobiology and Vaccine Research Laboratory, First Department of  
12 Pediatrics, "Aghia Sophia" Children's Hospital, School of Medicine, National  
13 and Kapodistrian University of Athens
- 14 3. 1st Department of Critical Care & Pulmonary Services, Medical School,  
15 National and Kapodistrian University of Athens, Greece
- 16 4. Unit of Bioinformatics and Applied Genomics, Department of Microbiology,  
17 Hellenic Pasteur Institute, Athens, Greece
- 18 5. Department of Zoology, University of Oxford, United Kingdom
- 19 6. Imperial College, London, United Kingdom
- 20 7. Birmingham City University, Birmingham, United Kingdom
- 21 8. National Public Health Organization, Athens, Greece
- 22 9. Department of Histology and Embryology, School of Medicine, National  
23 Kapodistrian University of Athens, Greece
- 24 10. Biomedical Research Foundation, Academy of Athens, Athens, Greece

25 11.Center for New Biotechnologies and Precision Medicine, Medical School,  
26 National and Kapodistrian University of Athens, Athens, Greece

27 12.Molecular and Clinical Cancer Sciences, Manchester Cancer Research  
28 Centre, Manchester Academic Health Sciences Centre, University of  
29 Manchester, Manchester, UK

30 13.4th Department of Internal Medicine, Attikon University Hospital,  
31 MedicalSchool, National and Kapodistrian University of Athens, Greece

32 \*Correspondence should be addressed to:

33 Gkikas Magiorkinis MD PhD FRCPATH

34 75 Mikras Asias Street, Greece

35 National Retrovirus Reference Centre, Department of Hygiene, Epidemiology and  
36 Medical Statistics, Medical School, National and Kapodistrian University of Athens

37 gmagi@med.uoa.gr

38

## 39 Abstract

40 Background: Severe COVID-19 pneumonia has been associated with the  
41 development of intense inflammatory responses during the course of infections with  
42 SARS-CoV-2. Given that Human Endogenous Retroviruses (HERVs) are known to  
43 be activated during and participate in inflammatory processes, we examined whether  
44 HERV dysregulation signatures are present in COVID-19 patients.

45 Results: By comparing transcriptomes of Peripheral Blood Monocytes (PBMCs) and  
46 Bronchoalveolar Lavage Fluid (BALF) from patients and normal controls we have  
47 shown that HERVs are intensely dysregulated in BALF, but not in PBMCs. In  
48 particular, upregulation in the expression of multiple HERV families was detected in  
49 BALF samples of COVID-19 patients, with HERV-W being the most highly  
50 upregulated family among the families analysed. In addition, we compared the  
51 expression of HERVs in Human Bronchial Epithelial Cells (HBEs) without and after  
52 senescence induction in an oncogene-induced senescence model, in order to  
53 quantitatively measure changes in the expression of HERVs in bronchial cells during  
54 the processes of cellular senescence.

55 Conclusions: This apparent difference of HERV dysregulation between PBMCs and  
56 BALF warrants further studies in involvement of HERVs in inflammatory  
57 pathogenetic mechanisms as well as exploration of HERVs as potential biomarkers  
58 for disease progression. Furthermore, the increase in the expression of HERVs in  
59 senescent HBEs in comparison to non-induced HBEs provides a potential link for  
60 increased COVID-19 severity and mortality in aged populations.

61

62 **Keywords:** COVID-19, endogenous retroviruses, HERV-W, inflammation,  
63 senescence

## 64     **Background**

65

66             Human Endogenous Retroviruses (HERVs) are the remnants of ancient  
67     retroviral infections that infiltrated the germlines of our deep in time ancestors and  
68     integrated within their genomes. Through Mendelian inheritance and evolutionary  
69     processes spanning millions of years they now occupy around 8% of the human  
70     genome (1). Their evolutionary history based on analyses of molecular sequences  
71     suggests that at least 30 independent colonization events generated an equivalent  
72     number of closely related retroviral integrations known as families (2).

73             The vast majority of HERV integrations have accumulated mutations that  
74     effectively incapacitated their proliferative functionality, however the these integrants  
75     can still be transcribed and encode proteins (3). HERV expression is usually silenced  
76     through numerous post-transcriptional mechanisms, but can be upregulated in  
77     certain diseases and conditions such as cancer and inflammation (4,5). The  
78     retroviral involvement in inflammatory processes has been shown to occur through  
79     over-production of nucleic acids and proteins that interfere with a variety of innate  
80     immune response and inflammatory pathways (6).

81             COVID-19 emerged in late 2019 in Wuhan City and from China spread  
82     around the globe generating the most intensive pandemic responses within the last  
83     50 years leading to significant socioeconomic disruption. The disease is caused by a  
84     zoonotic coronavirus, now known as SARS-CoV-2, which in a proportion of  
85     individuals will cause severe pneumonia, respiratory and multi-organ failure referred  
86     to as Corona virus infectious disease 2019 (COVID-19) (7). For severe COVID-19 it  
87     has been suggested that a critical component of the pathophysiology is a severe  
88     inflammatory response driven by cytokine release (8).

Here we aimed to explore whether HERV expression is upregulated in patients with COVID-19 as this could be consistent with an involvement of HERVs in inflammatory responses such as the production of interferonogenic nucleic acids and inflammatory proteins. In the initial study, in which the COVID-19 patients' RNAseq data were produced, the authors correlated the increased expression of proinflammatory cytokines and cytokine receptors in Bronchoalveolar Lavage Fluid (BALF) samples of COVID-19 patients when compared to healthy subjects, and related this finding to the cytokine storm and disease severity in these patients (9). In addition, we aimed to determine the differences between the expression of HERVs after oncogene-induced senescence of Human Bronchial Epithelial Cells (HBECS) in a non-malignant senescence model and the expression of HERVs in non-induced HBECS (10,11). In this way, we aim to determine the effect of senescence on the expression of HERVs as a potential component of the increased inflammatory burden of senescence in bronchial cells (10). We hypothesized that the expression of HERVs is augmented in the state of senescence and this could offer a plausible explanation for the increased COVID-19 severity and mortality observed as the patient age increases.

We find that multiple HERV families are upregulated in BALF, but not in Peripheral Blood Monocytes (PBMCs), in patients with COVID-19 compared to healthy individuals. The findings merit further study regarding the potential involvement of HERVs in COVID-19.

## Results

### Expression of HERVs in BALF samples and PBMCs from COVID-19 patients in comparison to healthy controls

Transcription of HERVs was normalized by means of two housekeeping genes, succinate dehydrogenase (SDHA) and hypoxanthine phosphoribosyl transferase 1 (HPRT1). Upregulation in the expression of endogenous retroviral elements was observed in the BALF samples of COVID-19 patients in comparison to healthy controls. In particular, in the normalized expression, with the SDHA expression, statistically significant upregulation in the BALF samples of COVID-19 patients in comparison to healthy controls was found in the expression of HERV-K (HML-1) (2.94 fold change,  $p=0.047$ ), HERV-K (HML-2) (2.59 fold change,  $p=0.041$ ), HERV-K (HML-3) (4.95 fold change,  $p<0.001$ ), HERV-K (HML-5) (4.5 fold change,  $p=0.002$ ), HERV-K (HML-6) (2 fold change,  $p=0.005$ ), HERV-W (8.33 fold change,  $p<0.001$ ), HERV-H (6.64 fold change,  $p<0.001$ ), ERV-L (6.33 fold change,  $p<0.001$ ), HERV-9 (5.44 fold change,  $p=0.004$ ), HERV-E (4.65 fold change,  $p<0.001$ ), HERV-FRD (4.35 fold change,  $p=0.009$ ) (Figure 1 and Figure 2). No statistically significant differences were observed in the expression of HERV-K (HML-4) ( $p=0.709$ ) and HERV-I ( $p=0.053$ ).

In the normalization with the HPRT1 expression, the upregulation in the expression of HERVs in COVID-19 patients was statistically significant in the expression of HERV-K (HML-3) (6.79 fold change,  $p=0.017$ ), HERV-K (HML-5) (6.41 fold change,  $p=0.033$ ), HERV-W (11.06 fold change,  $p=0.005$ ), HERV-H (8.73 fold change,  $p=0.006$ ), ERV-L (8.25 fold change,  $p=0.006$ ), HERV-9 (8.18 fold change,  $p=0.036$ ), HERV-E (6.05 fold change,  $p=0.007$ ), HERV-FRD (5.31 fold change,  $p=0.014$ ) and HERV-I (4.47 fold change,  $p=0.014$ ), while no statistically significant

upregulation was observed in the expression of HERV-K (HML-1), HERV-K (HML-2), HERV-K (HML-4), and HERV-K (HML-6) (Figure 3 and Figure 4).

In the normalization with the SDHA expression in the PBMCs datasets, statistically significant downregulation of the expression of ERV-L (0.85 fold change,  $p=0.029$ ) and HERV-FRD (0.64 fold change,  $p=0.049$ ) was observed in COVID-19 patients' PBMCs in comparison to PBMCs of healthy donors. No statistically significant differences were observed in the expression of HERV-K (HML-1), HERV-K (HML-2), HERV-K (HML-3), HERV-K (HML-4), HERV-K (HML-5), HERV-K (HML-6), HERV-W, HERV-H, HERV-9, HERV-E, and HERV-I (Figure 5 and Figure 6 in Additional Files 1 and 2 respectively).

In the normalized expression with the HPRT1 expression, statistically significant downregulation was observed in the expression of HERV-K (HML-1) in the PBMCs of COVID-19 patients in comparison to PBMCs of healthy controls (0.74 fold change,  $p=0.007$ ). No statistically significant differences were observed in the expression of HERV-K (HML-2), HERV-K (HML-3), HERV-K (HML-4), HERV-K (HML-5), HERV-K (HML-6), HERV-W, HERV-H, ERV-L, HERV-9, HERV-E, HERV-FRD and HERV-I (Figure 7 and Figure 8 in Additional Files 3 and 4 respectively).

### **Expression of HERVs in non-induced and induced (senescent) HBECs**

Transcription of HERVs was normalized by means of two housekeeping genes, succinate dehydrogenase (SDHA) and hypoxanthine phosphoribosyl transferase 1 (HPRT1). Upregulation in the expression of endogenous retroviral elements was recognised in the senescence induced HBECs in comparison to non-induced HBECs. After the normalization with the expression of SDHA gene, statistically significant upregulation in the induced HBECs was detected for HERV-K (HML-1) (1.91 fold change,  $p<0.001$ ), HERV-K (HML-2) (2.29 fold change,  $p<0.001$ ),

HERV-K (HML-3) (1.36 fold change,  $p=0.001$ ), HERV-K (HML-5) (1.67 fold change,  $p<0.001$ ), HERV-K (HML-6) (1.3 fold change,  $p=0.001$ ), HERV-W (1.9 fold change,  $p<0.001$ ), ERV-L (1.86 fold change,  $p<0.001$ ), HERV-9 (1.74 fold change,  $p<0.001$ ), HERV-E (3.38 fold change,  $p<0.001$ ). No statistically significant differences were observed in the expression of HERV-H, HERV-I and HERV-FRD.

After the normalization with the expression of HPRT1 gene, statistically significant upregulation in the induced HBECs was detected for HERV-K (HML-1) (1.57 fold change,  $p<0.001$ ), HERV-K (HML-2) (1.98 fold change,  $p<0.001$ ), HERV-K (HML-3) (1.16 fold change,  $p=0.047$ ), HERV-K (HML-5) (1.43 fold change,  $p<0.001$ ), HERV-W (1.61 fold change,  $p<0.001$ ), ERV-L (1.57 fold change,  $p<0.001$ ), HERV-9 (1.51 fold change,  $p<0.001$ ), HERV-E (3.57 fold change,  $p<0.001$ ). No statistically significant differences were observed in the expression of HERV-K (HML-6), HERV-H and HERV-FRD.

Downregulation in the expression of senescent HBECs in comparison to non-induced HBECs was recognised in the expression of HERV-K (HML-4) after the normalization with SDHA and HPRT1 gene (0.7 fold change,  $p=0.011$  and 0.61 fold change,  $p<0.001$  respectively) and in the expression of HERV-I after normalizing with the expression of HPRT1 gene (0.7 fold change,  $p<0.001$ ).



## Discussion

SARS-CoV-2 infection is either mild or asymptomatic in the majority of the infections (7,12). Age and underlying comorbidities like pulmonary and/or cardiovascular disease are major risk factors for severe COVID-19 (13). However, the underlying causes of the severity of the infection are not fully understood. Here we have explored whether HERVs are dysregulated in patients with COVID-19 as this could provide evidence in favour of the hypothesis that they could be implicated in the severity of the disease. Based on transcriptomic data we have found that multiple HERV families are upregulated in BALF, but not in PBMCs of COVID-19 patients. Our findings combined with data showing that entry factors for SARS-CoV-2 are co-expressed with innate immunity genes in respiratory cells suggests that HERV upregulation in BALF might indeed be relevant to local rather than systematic immunity responses (14).

While HERVs are not commonly expressed throughout life, they have on the other hand been shown to have a dual role with respect to inflammation. They are upregulated by inflammatory pathways (15), but also ERV proteins and nucleic acids may trigger inflammatory responses (16). For example, we have previously shown that IFI-16 has a broad range of ssDNA targets, thus if HERVs or other transposable elements are upregulated and reverse transcribed as a result of COVID-19, this could amplify inflammatory responses (17). In our analyses HERV-W was the most highly upregulated family (up to 11x times higher). Interestingly, HERV-W *env* has been shown to be associated with proinflammatory transcriptional signatures and induce proinflammatory responses (18). Whether the upregulation of HERV-W RNA seen in transcriptomes results in production of HERV-W proteins remains to be determined.

Crucially, HERV expression seems to have an association with age, which is the strongest risk factor for severe COVID-19. Differential expression of HERVs has been described when comparing older with younger adults (19). Upregulation of endogenous retroviruses as a result of aging has been studied in mice whereas failure of epigenetic mechanisms seems to be the driving force for upregulation of endogenous retroviruses with age (20). Hence, it could be hypothesized that a possible lack of epigenetic regulation of HERVs in older individuals, may result in HERV expression overdrive during COVID-19 which in turn results in massive pro-inflammatory responses (21). This association between aging and increased expression of endogenous retroviral elements is confirmed by our findings. In the frameworks of this study an upregulation in the expression of most HERV families included in our analyses is demonstrated in senescence induced HBECs in comparison to non-induced. These results indicate that the increased expression of endogenous retroviral elements in senescent bronchial epithelial cells possibly induces local chronic inflammation, as a hallmark of immunosenescence and enhance the hypothesis of local rather than systematic inflammatory reactions. This hypothesis provides a plausible explanation for the increased COVID-19 severity and mortality as age increases.

Furthermore, recently, COVID-19 exposure has been linked to hyperinflammatory shock in paediatric patients with manifestations that resemble Kawasaki Disease (both classic and incomplete), thus COVID-19 was linked to an unexpected increase of cases of the syndrome (22). Intriguingly, in a case series of Kawasaki paediatric patients linked to the COVID-19 epidemic, HERV levels were detectable in the microbiology results of one of the eight afflicted children described

in this case series (22). This finding could potentially indicate a further role of HERVs as an inflammatory trigger upon SARS-CoV-2 infection.

One limitation of our study is the small number of samples; thus, our findings need to be replicated in more patients, but also need to be followed up by functional studies. Nevertheless, our findings might also (at least partially) provide a working hypothesis to explain the mechanism of action of Lopinavir/ritonavir which has been shown to inhibit the expression of at least one HERV family (HK2) (23). Interestingly, there is also evidence of the potential benefit of antiretroviral treatment against SARS-CoV-2 infection, as there is an ongoing phase 2b/3 Clinical Trial for the assessment of the effectiveness of Ieronlimab, a CCR5 antagonist, against COVID-19 (24).

We find that multiple HERV families are upregulated in Bronchoalveolar Lavage Fluid (BALF), but not in PBMCs, in patients with COVID-19 compared to healthy individuals. Furthermore, we were able to identify upregulation in the expression of HERVs in senescence induced HBECs in comparison to non-induced cells, a fact that indicates the potential role of increased endogenous retroviral expression as a mediator of inflammatory reactions in older individuals that are at increased risk due to the disease. The findings merit further study regarding the potential involvement of HERVs in COVID-19. It thus seems feasible that should HERV expression be aetiologically linked to COVID-19 this expression could be a therapeutic target that to minimize the likelihood of severe Covid-19 and death. It remains to be seen if HERV expression could act as a potential marker of disease severity.

## Conclusions

In this study, we recognised dysregulated expression of endogenous retroviral elements in BALF samples, but not in PBMCs of COVID-19 patients. At the same time, we were able to identify upregulated expression of multiple HERV families in senescence induced HBECs in comparison to non-induced, a fact that could possibly explain the differences in disease severity among age groups. These results indicate that HERV expression might play a pathophysiological role in local inflammatory pathways in lungs afflicted by SARS-CoV-2 and their expression could be a potential therapeutic target.

## Methods

In order to compare the expression of endogenous retroviral elements in the BALF and PBMCs of COVID-19 patients and healthy individuals, online available data were utilized. RNA sequencing (RNAseq) data from BALF samples of 2 COVID-19 patients (2 samples from each) and PBMCs of 3 COVID-19 patients and 3 healthy donors were downloaded in .fastq format from National Genomics Data Center- Genome Sequence Archive (NGDC-GSA) with accession numbers: CRR119890, CRR119891, CRR119892, CRR119893, CRR119894, CRR119895, CRR119896, CRR119897, CRR125445, CRR125446. RNAseq data from BALF samples of healthy individuals were downloaded from Sequence Read Archive (SRA) with accession numbers: SRR10571724, SRR10571730, SRR10571732. All data were paired-end. The platforms used for the production of the raw data were BGISEQ-500 for CRR119890, CRR119891, CRR119892 and CRR119893 (reads of 100 bases length), Illumina MiSeq for the production of CRR119894, CRR119895, CRR119896 and CRR119897 (reads of 150 bases length), Illumina NovaSeq 5000 for the production of CRR125445 and CRR125446 (reads of 150 bases length), and finally Illumina HiSeq 2000 for the production of SRR10571724, SRR10571730 and SRR10571732 (reads of 50 bases length).

For the evaluation of the effect of senescence on the expression of HERV families, data from a non-malignant epithelial oncogene-induced senescence model was used in order to produce senescent HBECs as was described by Komseli et al (11). Non-induced and induced (senescent) HBECs online available RNAseq data were retrieved from NCBI-SRA with accession numbers: SRR6261633, SRR6261634, SRR6261635, SRR6261636, SRR6261637, SRR6261638, SRR6261639, SRR6261640, SRR6261641, SRR6261642, SRR6261643,

SRR6261644, SRR6261645, SRR6261646, SRR6261647, SRR6261648,  
SRR6261649, SRR6261650, SRR6261651, SRR6261652, SRR6261653,  
SRR6261654, SRR6261655, SRR6261656. For the production of these data Illumina  
NextSeq500, producing paired-end data (reads of 38 bases length).

### *Bioinformatics Analysis*

Bowtie2 with default settings for paired-end data was used for the alignment  
of these data to hg19 human genome assembly (25). Samtools view command with  
the -q option was used, in order for the mapping quality of the reads included in the  
alignments to be over 20 (26). Samtools sort and index commands were used with  
default settings. IGV was used for the visualisation of the mapping alignments (27).

For the identification of the expression of endogenous retroviral elements, the  
coordinates of HERV-K (HML-2), HERV-H, HERV-W, HERV-L, HERV-E, HERV-I,  
HERV-9, HERV-FRD, HERV-K (HML-1), HERV-K (HML-3), HERV-K (HML-4),  
HERV-K (HML-5) and HERV-K (HML-6) elements were found in the existing literature  
(28-31). HERV-K (HML-2), HERV-H and HERV-W elements' coordinates were set in  
reference genome hg19. Regarding the rest of the endogenous retroviral elements,  
their coordinates were referring to hg38 assembly and the UCSC Batch Coordinate  
Conversion (liftOver) online tool was used for the conversion to hg19 coordinates  
(32).

Bedtools multicov command was used for the quantification of endogenous  
retroviral expression in each of the datasets used (33). The -f option was used in  
order to reassure that at least 80% of the read length were overlapping with the  
HERVs elements' coordinates. For this reason, based on the coordinates we used  
and the length of endogenous retroviral elements, we considered a length of 9000bp

for HML-3, HML-5 and HML-6, 10000bp for HML-1, HML-2, HML-4 and HML-6, HERV-W, HERV-L, HERV-E, HERV-I and HERV-FRD, a length of 9000bp for HML-3, HML-5 and HML-6, 12000bp for HERV-9 and 17000bp for HERV-H for the calculation of -f for each virus, in order to increase the sensitivity of the detection.

Taking into account the different read lengths and different depths of sequencing among the datasets used in this analysis, we utilised the expression of widely used housekeeping genes for the normalization of our raw reads in each of the datasets. For this purpose, we calculated the expression of SDHA and HPRT1 genes in the datasets used in this analysis. Bedtools multicov command was used for the quantification of the expression of these genes with the -f option for the detection of reads that overlap with the genes coordinates by at least 80%, the same way as with the endogenous retroviral elements. Finally, the sum of the raw read counts from each ERV in each dataset was calculated and was normalized by dividing this number by the expression (in raw reads) of each of the housekeeping genes used, in the respective dataset (normalized expression value). The fold change in the expression of the endogenous retroviral elements in COVID-19 BALF samples and PBMCs was calculated as the ratio of the mean normalized expression value of ERVs in COVID-19 patients to the mean normalized expression value of ERVs in healthy individuals. Respectively, the fold change for the expression between non-induced and induced HBECs was calculated as the ratio of the mean normalized expression value of ERVs in induced (senescent) HBECs to the mean normalized expression value of ERVs in non-induced HBECs.

### *Statistical Analysis*

IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0.

Armonk, NY: IBM Corp. was used for the statistical analysis. The normalized expression value of each ERV was log-transformed ( $\ln[\text{normalized expression value}]$ ) in order to perform independent sample t-test, with unequal variances assumed (Welch's t-test), in order to identify statistically significant differences between the BALF samples and PBMCs of COVID-19 patients and healthy controls as well as to identify statistically significant differences in the expression of the studied HERVs between HBECs without and with oncogene-induced senescence.

### **List of Abbreviations**

COVID-19 (Corona Virus Disease 2019)

SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus)

HERVs (Human Endogenous Retroviruses)

PBMCs (Peripheral Blood Monocytes)

BALF (Bronchoalveolar Lavage Fluid)

HBECs (Human Bronchial Epithelial Cells)

SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2)

SDHA (Succinate Dehydrogenase)



Hypoxanthine Phosphoribosyl Transferase 1 (HPRT1)

HERV-W *env* (HERV-W envelope protein)

hg19 (human genome, version 19)

hg38 (human genome, version 38)

ERV (Endogenous retrovirus)

## Declarations

### ***Ethics Approval and Consent to Participate***

Not applicable

### ***Consent for Publication***

Not applicable

### ***Availability of Data and Materials***

The datasets of the PBMCs from COVID-19 patients and healthy volunteers, as well as the data of BALF from COVID-19 patients analysed during the current study are available in National Genomics Data Center-Genome Sequence Archive (NGDC-GSA), in BIG Data Center (<https://bigd.big.ac.cn/>), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences with accession numbers CRR119890, CRR119891, CRR119892, CRR119893, CRR119894, CRR119895, CRR119896,

CRR119897, CRR125445 and CRR125446  
(<https://bigd.big.ac.cn/gsa/browse/CRA002390>) (9).

The datasets of BALF from healthy controls analysed in the present work are available in NCBI-Sequence read Archive (SRA), under accession numbers: SRR10571724 (<https://www.ncbi.nlm.nih.gov/sra/?term=srr10571724>), SRR10571730 (<https://www.ncbi.nlm.nih.gov/sra/?term=srr10571730>), SRR10571732 (<https://www.ncbi.nlm.nih.gov/sra/?term=srr10571732>).

The datasets of the induced and non-induced HBECs analysed in the present work are available in NCBI-Sequence read Archive (SRA), under accession numbers: SRR6261633, SRR6261634, SRR6261635, SRR6261636 (<https://www.ncbi.nlm.nih.gov/sra?term=SRX3367915>), SRR6261637, SRR6261638, SRR6261639, SRR6261640 (<https://www.ncbi.nlm.nih.gov/sra?term=SRX3367916>), SRR6261641, SRR6261642, SRR6261643, SRR6261644 (<https://www.ncbi.nlm.nih.gov/sra?term=SRX3367917>), SRR6261645, SRR626163346, SRR6261647, SRR6261648 (<https://www.ncbi.nlm.nih.gov/sra?term=SRX3367918>), SRR6261649, SRR6261650, SRR6261651, SRR6261652 (<https://www.ncbi.nlm.nih.gov/sra?term=SRX3367919>), SRR6261653, SRR6261654, SRR6261655, SRR6261656 (<https://www.ncbi.nlm.nih.gov/sra?term=SRX3367920>).

### ***Competing Interests***

The authors declare that they have no competing interests.

### ***Funding***

Emblematic action to handle SARS-CoV-2 infection: Epidemiological study in

413 Greece via extensive testing for viral and antibody detection, sequencing  
414 of the virome and genetic analysis of the carriers (GSRT No 2020ΣE01300001).

415

# 416 **Acknowledgements**

417 The study is supported by:

418 GSRT No 2020ΣE01300001

419 Emblematic action to handle SARS-CoV-2 infection: Epidemiological study in

420 Greece via extensive testing for viral and antibody detection, sequencing

421 of the virome and genetic analysis of the carriers (GSRT No 2020ΣE01300001).

422

## References

1. Griffiths DJ. Endogenous retroviruses in the human genome sequence. *Genome Biol* 2001;2:REVIEWS1017.
2. De Parseval N, Heidmann T. Human endogenous retroviruses: From infectious elements to human genes. *Cytogenet Genome Res*. 2005;110:318-32.
3. Chen J, Foroozesh M, Qin Z. Transactivation of human endogenous retroviruses by tumor viruses and their functions in virus-associated malignancies. *Oncogenesis*. 2019;8:6.
4. Johanning GL, Malouf GG, Zheng X, Esteva FJ, Weinstein JN, Wang-Johanning F, et al. Expression of human endogenous retrovirus-K is strongly associated with the basal-like breast cancer phenotype. *Sci Rep*. 2017;7:41960.
5. Greenig M. HERVs, immunity, and autoimmunity: Understanding the connection. *PeerJ*. 2019;7:e6711.
6. Grandi N, Tramontano E. Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. *Front Immunol*. 2018;9:2039.
7. Wu Z, McGoogan JM. Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention. *JAMA*. 2020. doi: 10.1001/jama.2020.2648. [Epub ahead of print] Available from: <https://jamanetwork.com/journals/jama/fullarticle/2762130>
8. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis.

Clin Chem Lab Med. 2020. doi: 10.1515/cclm-2020-0369. [Epub ahead of print]. Available from: <https://www.degruyter.com/view/journals/cclm/ahead-of-print/article-10.1515-cclm-2020-0369/article-10.1515-cclm-2020-0369.xml>

9. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect.* 2020;9:761-770.

10. Gorgoulis V, Adams PD, Alimonti A, Bennett DC, Bischof O, Bishop C, et al. Cellular Senescence: Defining a Path Forward. *Cell.* 2019 ;179:813-827.

11. Komseli ES, Pateras IS, Krejsgaard T, Stawiski K, Rizou S V., Polyzos A, et al. A prototypical non-malignant epithelial model to study genome dynamics and concurrently monitor micro-RNAs and proteins in situ during oncogene-induced senescence. *BMC Genomics.* 2018;19:37.

12. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* 2020;25.

13. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395:1054-1062.

14. Sungnak W, Huang N, Bécavin C, Berg M, Queen R, Litvinukova M, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med.* 2020. doi: 10.1038/s41591-020-0868-6. [Epub ahead of print]. Available from: <http://www.nature.com/articles/s41591-020-0868-6>

15. Manghera M, Ferguson-Parry J, Lin R, Douville RN. NF- $\kappa$ B and IRF1 Induce Endogenous Retrovirus K Expression via Interferon-Stimulated Response Elements in Its 5' Long Terminal Repeat. *J Virol*. 2016;90:9338–49.
16. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell*. 2015;162:974–86.
17. Hurst TP, Aswad A, Karamitros T, Katzourakis A, Smith AL, Magiorkinis G. Interferon-inducible protein 16 (IFI16) has a broad-spectrum binding ability against ss DNA targets: An evolutionary hypothesis for antiretroviral checkpoint. *Front Microbiol*. 2019;10:1426.
18. Bhat RK, Ellestad KK, Wheatley BM, Warren R, Holt RA, Power C. Age- and disease-dependent HERV-W envelope allelic variation in brain: Association with neuroimmune gene expression. *PLoS One*. 2011;6:e19176.
19. Nevalainen T, Autio A, Mishra BH, Marttila S, Jylhä M, Hurme M. Aging-associated patterns in the expression of human endogenous retroviruses. *PLoS One*. 2018;13:e0207407.
20. Crichton JH, Dunican DS, MacLennan M, Meehan RR, Adams IR. Defending the genome from the enemy within: Mechanisms of retrotransposon suppression in the mouse germline. *Cell Mol Life Sci*. 2014;71:1581-605.
21. Hurst TP, Magiorkinis G. Epigenetic control of human endogenous retrovirus expression: Focus on regulation of long-terminal repeats (LTRs). *Viruses*. 2017;9. pii: E130.
22. Riphagen S, Gomez X, Gonzalez-Martinez C, Wilkinson N, Theocharis P. Hyperinflammatory shock in children during COVID-19 pandemic. *Lancet*. 2020. DOI:[https://doi.org/10.1016/S0140-6736\(20\)31094-1](https://doi.org/10.1016/S0140-6736(20)31094-1) [Epub ahead of

498 print] Available from:  
499 <https://linkinghub.elsevier.com/retrieve/pii/S0140673620310941>  
500 23. Tyagi R, Li W, Parades D, Bianchet MA, Nath A. Inhibition of human  
501 endogenous retrovirus-K by antiretroviral drugs. *Retrovirology*. 2017;14:21.  
502 24. First Patient Treated with Leronlimab in Phase 2b/3 Trial for COVID-19::  
503 CytoDyn Inc. (CYDY) [Internet]. [cited 2020 May 8]. Available from:  
504 [https://www.cytodyn.com/newsroom/press-releases/detail/417/first-patient-](https://www.cytodyn.com/newsroom/press-releases/detail/417/first-patient-treated-with-leronlimab-in-phase-2b3-trial)  
505 [treated-with-leronlimab-in-phase-2b3-trial](https://www.cytodyn.com/newsroom/press-releases/detail/417/first-patient-treated-with-leronlimab-in-phase-2b3-trial)  
506 25. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nature*  
507 *Methods*. 2012;9:357-359.  
508 26. Li H.\*, Handsaker B.\*, Wysoker A., Fennell T., Ruan J., Homer N. et al. The  
509 Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*.  
510 2009;25:2078-9.  
511 27. James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman,  
512 Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer.  
513 28. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification,  
514 characterization, and comparative genomic distribution of the HERV-K (HML-  
515 2) group of human endogenous retroviruses. *Retrovirology*. 2011;8:90.  
516 29. Wang J, Xie G, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, et al. Primate-  
517 specific endogenous retrovirus-driven transcription defines naive-like stem  
518 cells. *Nature*. 2014;516:405–9.  
519 30. Grandi N, Cadeddu M, Blomberg J, Tramontano E. Contribution of type W  
520 human endogenous retroviruses to the human genome: Characterization of  
521 HERV-W proviral insertions and processed pseudogenes. *Retrovirology*.  
522 2016;13:1–25.

31. Bendall ML, De Mulder M, Iñiguez LP, Lecanda-Sánchez A, Pérez-Losada M, Ostrowski MA, et al. Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. PLoS Comput Biol. 2019;15:e1006453.
32. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002;12:996-1006. UCSC Genome Browser Utilities, Batch Coordinate Conversion (liftOver) (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>)
33. Quinlan AR and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–842.

## Figures, tables and additional files

### Figure 1:

Title: Upregulation of HML-1, HML-2, HML-3, HML-5, HML-6 in BALF of COVID-19 patients (SDHA normalization)

Legend: : Expression of HERV-K (HML-1), HERV-K (HML-2), HERV-K (HML-3), HERV-K (HML-4), HERV-K (HML-5), HERV-K (HML-6) in BALF of healthy controls and COVID-19 patients normalized with the expression of SDHA gene. Statistically significant upregulation in the BALF samples of COVID-19 patients in comparison to healthy controls was observed in the expression of HML-1 (2.94 fold change,  $p=0.047$ ), HML-2 (2.59 fold change,  $p=0.041$ ), HML-3 (4.95 fold change,  $p<0.001$ ), HML-5 (4.5 fold change,  $p=0.002$ ), HML-6 (2 fold change,  $p=0.005$ ).

### Figure 2:



Title: Upregulation of HERV-W, HERV-H, ERV-L, HERV-9, HERV-E, HERV-FRD the BALF of COVID-19 patients (SDHA normalization)

Legend: Expression of HERV-W, HERV-H, ERV-L, HERV-9, HERV-E, HERV-I and HERV-FRD in BALF of healthy controls and COVID-19 patients normalized with the expression of SDHA gene. Statistically significant upregulation in the BALF samples of COVID-19 patients in comparison to healthy controls was observed in the expression of HERV-W (8.33 fold change,  $p < 0.001$ ), HERV-H (6.64 fold change,  $p < 0.001$ ), ERV-L (6.33 fold change,  $p < 0.001$ ), HERV-9 (5.44 fold change,  $p = 0.004$ ), HERV-E (4.65 fold change,  $p < 0.001$ ), HERV-FRD (4.35 fold change,  $p = 0.009$ ).

### Figure 3:

Title: Upregulation of HML-3 and HML-5 in the BALF of COVID-19 patients (HPRT1 normalization)

Legend: Expression of HERV-K (HML-1), HERV-K (HML-2), HERV-K (HML-3), HERV-K (HML-4), HERV-K (HML-5), HERV-K (HML-6) in BALF of healthy controls and COVID-19 patients normalized with the expression of HPRT1 gene. Statistically significant upregulation in the BALF samples of COVID-19 patients in comparison to healthy controls was observed in the expression of HML-3 (6.79 fold change,  $p = 0.017$ ), HML-5 (6.41 fold change,  $p = 0.033$ ).

### Figure 4:

Title: Upregulation of HERV-W, HERV-H, ERV-L, HERV-9, HERV-E, HERV-FRD and HERV-I in BALF of COVID-19 patients (HPRT1 normalization)

Legend: Expression of HERV-W, HERV-H, HERV-L, HERV-9, HERV-E, HERV-I and

HERV-FRD in BALF of healthy controls and COVID-19 patients normalized with the expression of HPRT1 gene. Statistically significant upregulation in the BALF samples of COVID-19 patients in comparison to healthy controls was observed in the expression of HERV-W (11.06 fold change,  $p=0.005$ ), HERV-H (8.73 fold change,  $p=0.006$ ), ERV-L (8.25 fold change,  $p=0.006$ ), HERV-9 (8.18 fold change,  $p=0.036$ ), HERV-E (6.05 fold change,  $p=0.007$ ), HERV-FRD (5.31 fold change,  $p=0.014$ ) and HERV-I (4.47 fold change,  $p=0.014$ )

Additional File 1: File Format: Tagged Image File Format (.tif)

Title of Data: Figure 5: No statistically significant differences were observed in the expression of HERV-K (HML-1, HML-2, HML-3, HML-4 HML-5 and HML-6) between healthy donors and COVID-19 patients (SDHA normalization).

Description of Data: Expression of HERV-K (HML-1), HERV-K (HML-2), HERV-K (HML-3), HERV-K (HML-4), HERV-K (HML-5), HERV-K (HML-6) in PBMCs of healthy controls and COVID-19 patients normalized with the expression of SDHA gene. No statistically significant differences were observed between healthy donors and COVID-19 patients.

Additional File 2: File Format: Tagged Image File Format (.tif)

Title of Data: Figure 6: Downregulation of the expression of ERV-L and HERV-FRD in the PBMCs of COVID-19 patients (SDHA normalization).

Description of Data: Expression of HERV-W, HERV-H, HERV-L, HERV-9, HERV-E, HERV-FRD and HERV-I in PBMCs of healthy controls and COVID-19 patients normalized with the expression of SDHA gene. Statistically significant downregulation of the expression of ERV-L (0.85 fold change,  $p=0.029$ ) and

HERV-FRD (0.64 fold change,  $p=0.049$ ) was observed in the PBMCs of COVID-19 patients in comparison to PBMCs of healthy donors.

Additional File 3: File Format: Tagged Image File Format (.tif)

Title of Data: Figure 7: Downregulation in the expression of HML-1 in the PBMCs of COVID-19 patients (HPRT1 normalization).

Description of Data: Expression of HERV-K (HML-1), HERV-K (HML-2), HERV-K (HML-3), HERV-K (HML-4), HERV-K (HML-5), HERV-K (HML-6) in PBMCs of healthy controls and COVID-19 patients normalized with the expression of HPRT1 gene. Statistically significant downregulation was observed in the expression of HML-1 (0.74 fold change,  $p=0.007$ ) in the PBMCs of COVID-19 patients in comparison to PBMCs of healthy donors.

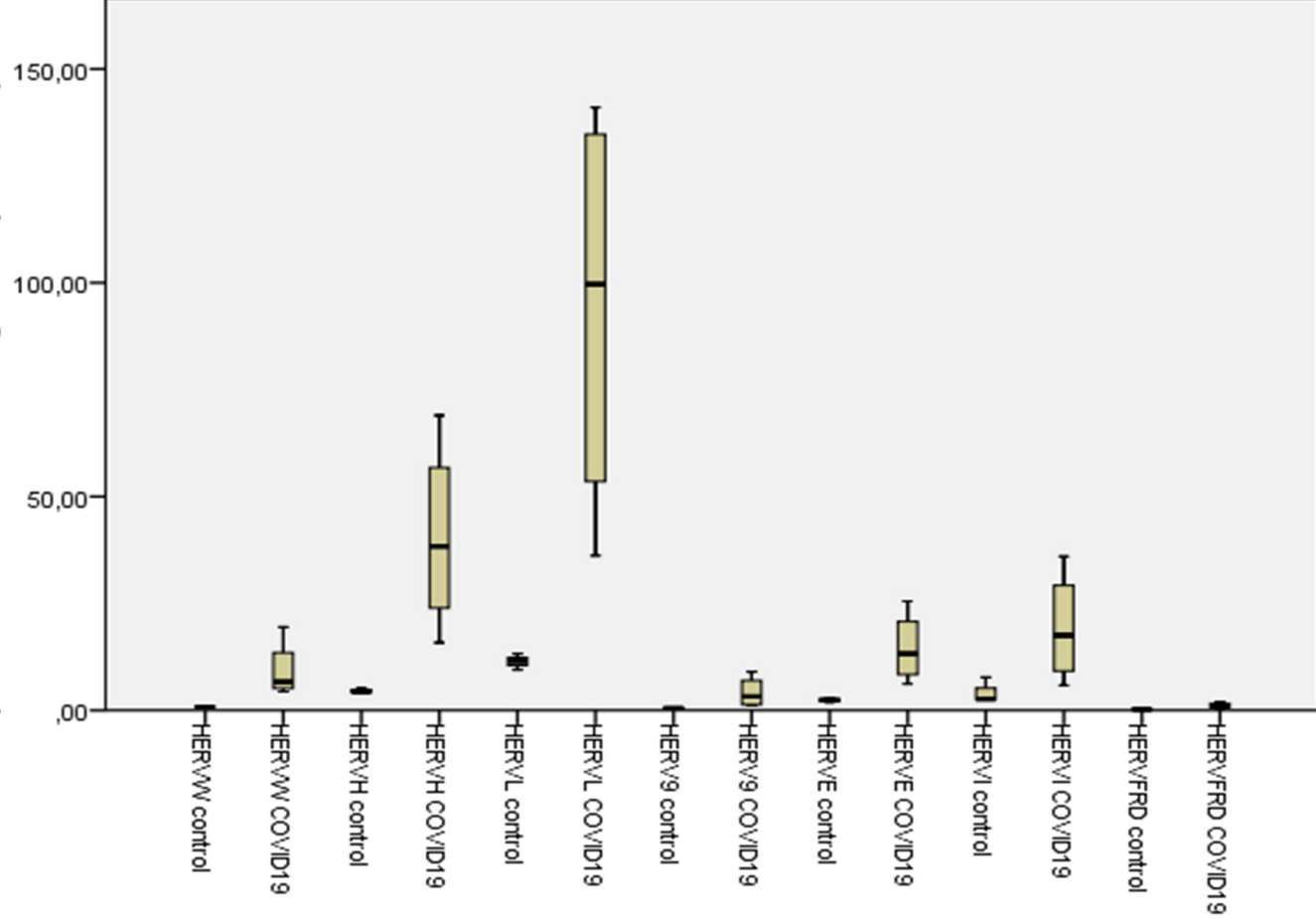
Additional File 4: File Format: Tagged Image File Format (.tif)

Title of Data: Figure 8: No statistically significant differences were observed in the expression of HERVs between healthy donors and COVID-19 patients.

Description of Data: Expression of HERV-W, HERV-H, HERV-L, HERV-9, HERV-E, HERV-FRD and HERV-I in PBMCs of healthy controls and COVID19 patients normalized with the expression of HPRT1 gene. No statistically significant differences were observed between healthy donors and COVID-19 patients.

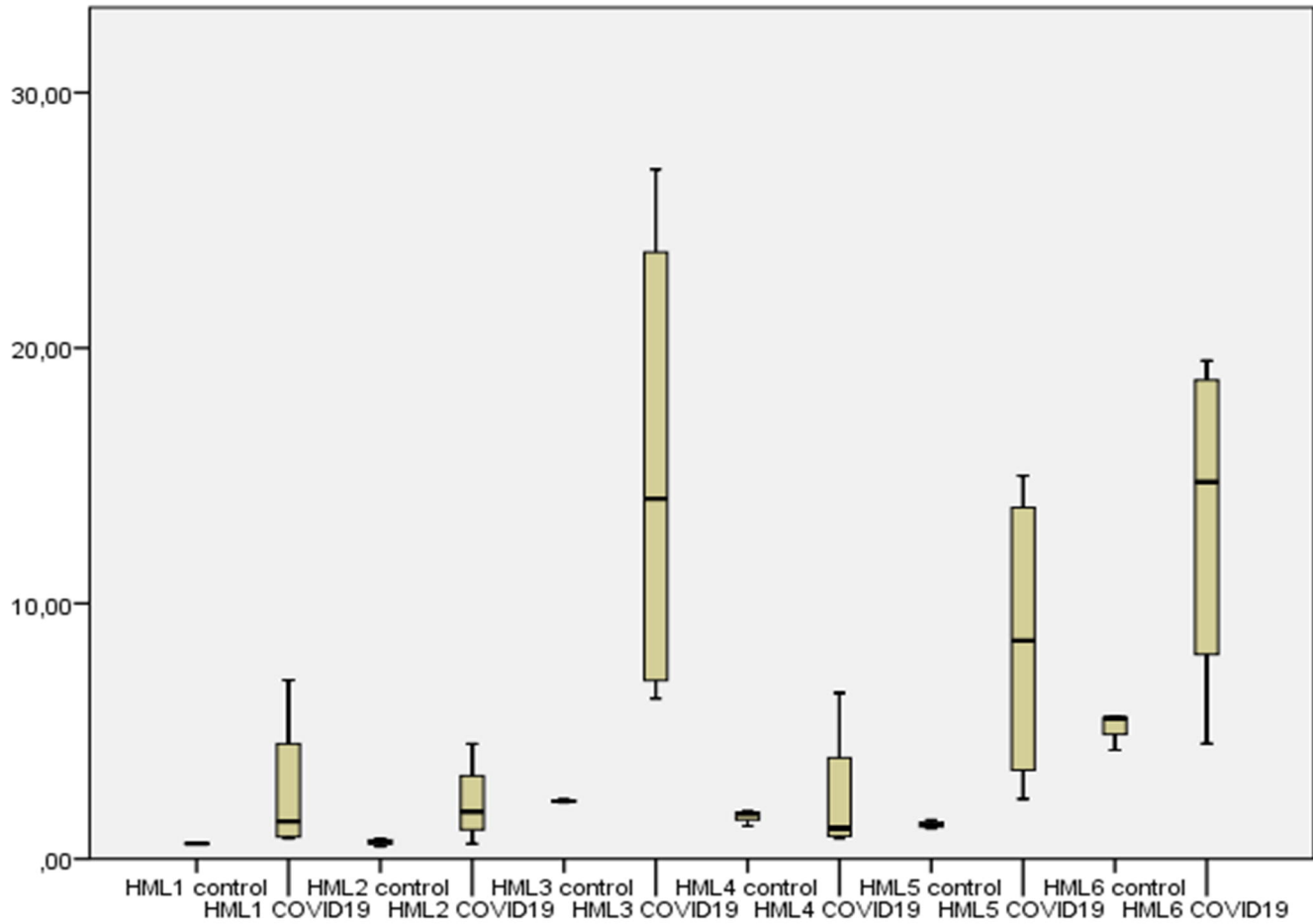
Expression in BALF samples

(normalized with HPRT1 gene expression)



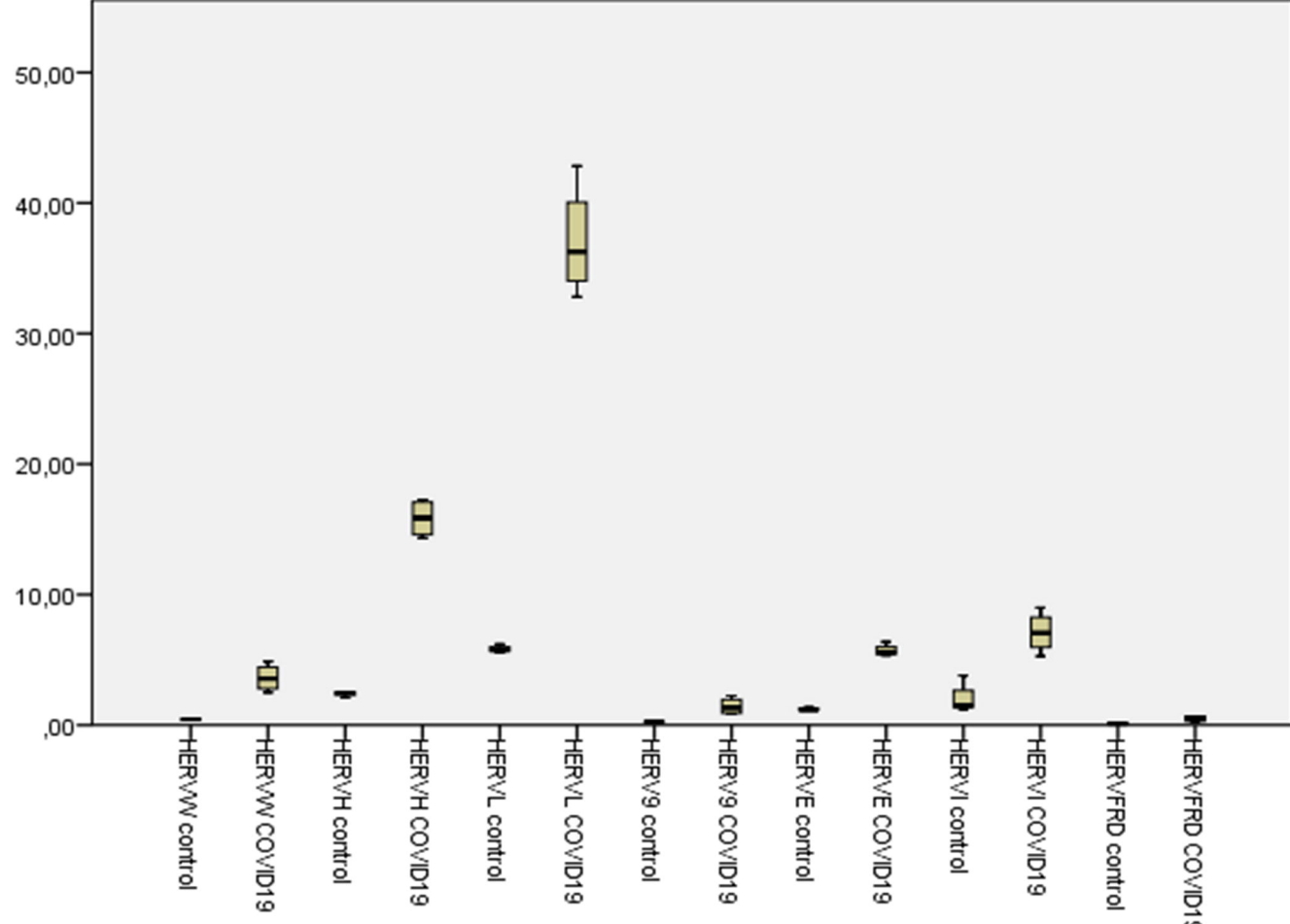
Expression in BALF samples

(normalized with HPRT1 gene expression)



Expression in BALF samples

(normalized with SDHA gene expression)



Expression in BALF samples  
(normalized with SDHA gene expression)

