



OPEN

# Blood leukocyte transcriptional modules and differentially expressed genes associated with disease severity and age in COVID-19 patients

Silvia Y. Bando<sup>1,8</sup>, Fernanda B. Bertonha<sup>1,8</sup>, Sandra E. Vieira<sup>1,8</sup>, Danielle B. L. de Oliveira<sup>2,3</sup>, Vanessa N. Chalup<sup>3</sup>, Edison L. Durigon<sup>3</sup>, Patricia Palmeira<sup>1</sup>, Ana Cristina P. Curi<sup>1</sup>, Caroline S. Faria<sup>4</sup>, Leila Antonangelo<sup>4</sup>, Gerhard da P. Lauterbach<sup>5</sup>, Fabiane A. Regalio<sup>6</sup>, Roberto M. Cesar Jr<sup>7</sup> & Carlos A. Moreira-Filho<sup>1</sup>✉

Since the molecular mechanisms determining COVID-19 severity are not yet well understood, there is a demand for biomarkers derived from comparative transcriptome analyses of mild and severe cases, combined with patients' clinico-demographic and laboratory data. Here the transcriptomic response of human leukocytes to SARS-CoV-2 infection was investigated by focusing on the differences between mild and severe cases and between age subgroups (younger and older adults). Three transcriptional modules correlated with these traits were functionally characterized, as well as 23 differentially expressed genes (DEGs) associated to disease severity. One module, correlated with severe cases and older patients, had an overrepresentation of genes involved in innate immune response and in neutrophil activation, whereas two other modules, correlated with disease severity and younger patients, harbored genes involved in the innate immune response to viral infections, and in the regulation of this response. This transcriptomic mechanism could be related to the better outcome observed in younger COVID-19 patients. The DEGs, all hyper-expressed in the group of severe cases, were mostly involved in neutrophil activation and in the p53 pathway, therefore related to inflammation and lymphopenia. These biomarkers may be useful for getting a better stratification of risk factors in COVID-19.

The molecular mechanisms determining COVID-19 severity in different age groups are not yet well understood. The study of blood transcriptomic differences between mild and severe cases of the disease seems to be a promising strategy for discovering these mechanisms, moreover when the information derived from the genomic analyses is combined with clinico-demographic and laboratory data obtained from patients' cohorts, in a systems biology approach<sup>1,2</sup>. Additionally, the obtention of gene expression profiles for severe and mild cases, and for patients' age groups as well, together with the identification and functional characterization of differentially expressed genes, may contribute to a better stratification of risk factors. It may also provide a thorough understanding of COVID-19 pathogenesis, helping the choice of adequate therapies<sup>3-5</sup> and, through the deciphering of host response transcriptional complexity, the discovery of new therapeutic targets<sup>5</sup>.

<sup>1</sup>Department of Pediatrics, Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP 05403-900, Brazil. <sup>2</sup>Instituto Israelita de Ensino e Pesquisa Albert Einstein, Hospital Israelita Albert Einstein, São Paulo, SP 01310-200, Brazil. <sup>3</sup>Department of Microbiology, Laboratory of Clinical and Molecular Virology, Institute of Biomedical Sciences, Universidade de São Paulo, São Paulo, SP 05508-040, Brazil. <sup>4</sup>Laboratório de Investigação Médica (LIM03), Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP 01246-903, Brazil. <sup>5</sup>Department of Internal Medicine, Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP 01246-903, Brazil. <sup>6</sup>Divisão de Anestesia, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP 01246-903, Brazil. <sup>7</sup>Department of Computer Science, Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, SP 05508-040, Brazil. <sup>8</sup>These authors contributed equally: Silvia Y. Bando, Fernanda B. Bertonha and Sandra E. Vieira. ✉email: cmoreira@usp.br

Consequently, we decided to investigate the transcriptional basis of the differences between severe and mild cases of COVID-19 through a comparative study of the transcriptional responses of human leukocytes to SARS-CoV-2 infection in different age groups. We adopted a Weighted Gene Co-expression Network Analysis (WGCNA) approach<sup>6</sup> for identifying transcriptional modules associated with the traits of interest (severity, age) and, subsequently, we conducted differential gene expression and enrichment analyses for discovering transcriptional biomarkers, as further described in the following paragraphs.

## Materials and methods

**Ethics statement.** This study was approved by the Research Ethics Committee of the Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) under number 4.001.109. A written informed consent was obtained from all participants or from their legal guardians. All methods were performed in accordance with the relevant guidelines and regulations.

**Characteristics of participants.** The participants included 121 SARS-CoV-2 PCR positive subjects recruited between May and August 2020, before the emergence of the first variant of concern in Brazil, that occurred in November 2020<sup>7</sup> (Supplementary Table S1). All these patients were unvaccinated against SARS-CoV-2. They were divided into two groups—Mild or Severe—according to the severity of illness categories described in the NIH COVID-19 Guidelines [COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health. Available at <https://www.covid19treatmentguidelines.nih.gov/>, accessed on 12/07/21]. The Severe group comprised 58 COVID-19 hospitalized patients requiring oxygen therapy that have been tested positive for SARS-CoV-2 by RT-qPCR and were admitted at the HC-FMUSP. The Mild group comprised 63 individuals recruited as outpatients who presented at least one of the following symptoms: fever, coryza, dyspnea, anosmia or hyposmia, ageusia or hypogeusia, wheezing in the chest, diarrhea, vomiting, body pain, headache, sore throat, or chills (Table 1). They were tested by nasopharyngeal swab RT-qPCR and found to be positive for SARS-CoV-2 and negative for other 15 respiratory viruses. After 14 days the clinical evolution of signs or symptoms of all these individuals was checked and the diagnosis of oligosymptomatic COVID-19 was confirmed.

**Factor analysis of mixed data (FAMD).** FAMD is a principal component method for analyzing a data set containing both numerical and categorical variables<sup>8</sup>. To perform this analysis, the Severe and Mild groups were divided into three age ranges: (i) 0 months to nine years; (ii) 10–40 years; (iii) > 41 years. Numerical data was obtained from Complete Blood Count (CBC) and age, while the qualitative levels of CBC variables (low, normal, or high) according to age and sex were included as categorical variables together with sex, symptoms, and other variables of potential clinical relevance (Table 1). This analysis excluded missing values and comprised 64 variables and 92 individuals (Supplementary Fig. S1). FAMD was performed using the package FactoMineR version 2.7<sup>9</sup> in R-studio environment (version 1.2.5033)<sup>10</sup>.

**Sample collection, processing, and analysis.** Whole blood samples were collected from hospitalized patients (median of 13 days after first symptoms) and from oligosymptomatic individuals (median of 10 days after first symptoms). All samples were collected in EDTA-containing tubes. Each blood sample was divided for complete blood count analysis, white blood cell and plasma separation. White blood cells were used for RNA extraction. Neutrophil-to-lymphocyte ratio (NLR), platelets-to-lymphocyte ratio (PLR), systemic immune-inflammation index ( $SII = (P \times N)/L$ ), and neutrophil-to-platelet ratio (NPR) were calculated as these hemogram-derived ratios correlate with inflammation and COVID-19 severity<sup>11</sup>.

Nasopharyngeal and oropharyngeal samples were collected through a swab and/or a mucus specimen trap and kept under refrigeration (4 °C) up to 24 h until cryogenic cooling and storage at –70 °C for further molecular diagnosis of SARS-CoV-2 and/or other respiratory viruses.

**Plasma cytokine concentrations.** Cytokines (IL-2, IL-4, IL-6, IL-10, IL-17, IFN-gamma, and TNF-alpha) were measured in plasma samples by flow cytometry using a BD™ Cytometric Bead Array (CBA) Human Th1/Th2/Th17 Cytokine Kit (BD Biosciences, San Jose, CA) according to the manufacturer's instructions, and the concentrations were expressed in pg/mL. Plasmatic IFN-alpha and IFN-beta were evaluated by ELISA (IFN beta and IFN alpha DuoSet ELISA, R&D Systems Inc., Minneapolis, MN) and the concentrations were expressed in pg/mL. Cytokine data were censored if they were below the detection limit of the instrument<sup>12</sup>.

**Detection of respiratory viruses by RT-qPCR.** Nasopharyngeal and oropharyngeal samples from oligosymptomatic individuals were obtained for RNA extraction using the NucliSens easyMag® platform fully automated (BioMerieux, Lyon, France), according to the manufacturer's instructions. RT-qPCR was employed for detecting the genetic material of SARS-CoV-2 and of 15 other respiratory viruses: influenza A virus (Inf A), influenza B virus (Inf B), seasonal coronaviruses (CoV-NL63, -229E, -HKU1, and -OC43), enterovirus (EV), parainfluenza viruses (PIV-1, -2, -3 and -4), human metapneumovirus (HMPV), rhinovirus (RV), respiratory syncytial virus (RSV), and adenovirus (AdV). A panel of validated in-house singleplex qPCR assays developed at the Centers for Disease Control and Prevention (CDC, Atlanta, GA, USA) was used according to Corman et al.<sup>13</sup> for SARS-CoV-2, and according to Sakthivel et al.<sup>14</sup> for the other viruses, using TaqMan™ assays (Applied Biosystems). For the reactions (25 µL of final volume), the AgPath-ID™ One-Step RT-PCR Kit (Applied Biosystems) was used, and the amplification step was carried out on the ABI 7500 instrument (Applied Biosystems) under the following

	Severe	Mild	<i>p</i> -value
Number of subjects	58	63	
Male	35 (60%)	25 (40%)	<b>0.012</b>
Median age	50 yrs	34 yrs	<b>0.002</b>
Male (range)	49 yrs (1 mo–81.4 yrs)	35 yrs (8 mo–79.8 yrs)	0.055
Female (range)	51 yrs (0 mo–75 yrs)	32 yrs (2.3–73.2 yrs)	<b>0.020</b>
0–40 yrs			
Male	13	15	
Female	6	26	<b>0.022*</b>
> 40 yrs			
Male	22	10	
Female	17	12	0.411*
Group 0–9 yrs (median)	0	5	<b>0.015</b>
Group 10–40 yrs (median)	36	29.5	<b>0.041</b>
Group > 40 yrs (median)	56.5	53	0.083
Race (no. subjects)			
White	37	37	0.302**
Mixed (Pardo)	21	11	0.140**
Asian	0	2	0.496**
Symptoms (no. subjects)			
Cough	Yes, No; U	Yes, No; U	
Cough	47; 11	45; 11; 5	0.325
Fever	42; 16	30; 28; 5	<b>0.011</b>
Coryza	12; 46	46; 12	<b>0.000</b>
Dyspnoea	42; 12	17; 41; 5	<b>0.000</b>
Wheezing in the chest	17; 40; 1	2; 56; 5	<b>0.000</b>
Anosmia/hyposmia	30; 23; 5	32; 24; 7	0.478
Ageusia/hypogeusia	27; 26; 5	32; 24; 7	0.260
Headache	25; 28; 5	8; 50; 5	<b>0.000</b>
Sore throat	34; 19; 5	31; 27; 5	0.128
Diarrhea	25; 33	17; 41; 5	0.062
Vomiting	15; 41; 2	3; 55; 5	<b>0.000</b>
Body pain	21; 31; 6	14; 43; 6	0.047
Chills	33; 23; 2	0; 44; 19	<b>0.027</b>
Therapy (no. subjects)			
Oxygen	49; 4; 5	NA	
Antibiotics	42; 11; 5	9; 49; 5	<b>0.000*</b>
Antiviral drugs	12; 41; 5	1; 57; 5	<b>0.001*</b>
Corticosteroids	19; 34; 5	3; 55; 5	<b>0.000*</b>
Hospital stay/age groups			
Median hospital stay (days)			
1 mo–9 yrs	7		0.311 (vs 10–40 yrs)
			0.045 (vs > 41 yrs)
10–40 yrs	6		<b>0.033</b> (vs > 41 yrs)
> 41 yrs	9		
Blood sample collection after symptoms onset (range)			
	Median 13 days (2–69 d)	Median 10 days (4–27 d)	<b>&lt; 0.0001</b>

**Table 1.** Clinical and demographic characteristics of the COVID-19 hospitalized (Severe group) and oligosymptomatic (Mild group) patients. *U* unknown, *NA* not applicable, *mo* months, *yrs* years, *d* days. Thirteen patients in the Mild group opted for not answering the question relative to race. All statistical analyses were performed using Mann Whitney test, except for the analyses indicated by \* or \*\* where Chi-square or Fisher exact tests were used. In bold, significant *p*-values < 0.05.

cycling conditions: 45 °C for 10 min (1 cycle); 95 °C for 10 min (1 cycle); and 95 °C for 15 s, followed by 55 °C for 1 min (45 cycles).

**RNA extraction.** A total of 0.5–1.5 mL of whole blood was used for white blood cells (WBC) RNA extraction. WBC were immediately separated by centrifugation using EL buffer (QIAamp RNA Blood Mini kit, Qiagen, Hilden, Germany). After cell separation, the WBC were collected, preserved in *RNAlater* (Qiagen) and stored at

10 °C until RNA extraction. WBC were lysed with RLT buffer, and the total RNA was extracted using QIAamp® RNA Blood Mini kit (Qiagen). RNA purity analysis and quantification were performed using the NanoVue spectrophotometer (GE Healthcare Life Sciences, Marlborough, MA). RNA quality was assessed on the Agilent BioAnalyzer 2100 (Agilent, Santa Clara, USA). All RNA samples were stored at –80 °C until used in hybridization experiments.

**Microarray hybridization.** A total of 23 RNA samples were used for gene expression analysis and grouped as Severe (n = 11) or Mild (n = 12) according to the patient's characteristics (Supplementary Table S2). These two groups showed no significant differences regarding age, sex, and time of whole blood collection after the first day of COVID-19 symptoms. The Severe group was further divided into subgroups A and B, where A included the samples from adolescents and younger adults (age ranging from 11 to 38 years; n = 5) and B included the samples from older adults (age ranging from 41 to 62 years; n = 6). Similarly, the Mild group was divided in subgroups C and D, where C included the samples from adolescents and younger adults (age ranging from 10 to 37 years; n = 7) and D included the samples from older adults (age ranging from 41 to 64 years; n = 5). The patients between zero and nine years (Severe and Mild groups) were not included due to insufficient RNA quality.

To determine gene expression profiles, 4 × 44 K DNA microarrays (Whole Human Genome Microarray Kit, Agilent Technologies, cat no. G4845A) were used. The procedures for hybridization using the fluorescent dye Cy3 followed the manufacturer's protocols (One-Color Microarray-Based Gene Expression Analysis—Quick Amp Labeling). The images were captured by the reader Agilent Bundle according to the parameters recommended for bioarrays and extracted by Agilent Feature Extraction software version 11.5.1.1 (<https://www.agilent.com/>). Spots with two or more flags (low intensity, saturation, controls, etc.) were considered as NA, that is, without valid expression value. All microarray raw data have been deposited in GEO public database (<http://www.ncbi.nlm.nih.gov/geo/>), a MIAME compliant database, under accession number GSE193022.

**Gene expression analysis.** An in-house algorithm in R environment (version 3.6.2)<sup>10</sup> was used for excluding transcripts presenting one or more missing values (NAs) per group and for converting gene expression values to log base 2. Through this procedure we obtained two gene expression data matrices: (i) one for gene co-expression network (GCN) analysis, and (ii) another for differential gene expression (DEG) analysis. The GCN matrix had 6375 Gene Ontology (GO) annotated genes after excluding all NAs. The DEG matrix had 15,248 GO annotated genes, including NAs and with a minimum of four valid gene expression values per group. Boxplot analysis was used for outlier detection (Supplementary Fig. S2). Data normalization for both data matrices was performed using limma package version 3.9<sup>15</sup> in R environment (version 3.6.2)<sup>10</sup>. The differential gene expression analyses were conducted in three comparisons separately: (i) Severe vs. Mild (ii) A vs. B; (iii) C vs. D.

**Weighted gene co-expression network analysis (WGCNA).** The network was constructed using the WGCNA package<sup>16</sup> (version 1.69-81; <https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/>) in R version 3.6.2 environment<sup>10</sup>. No outliers were detected in the analysis of gene expression data matrix. Pearson's correlation coefficient was used for obtaining gene co-expression similarity measures and for the subsequent construction of an adjacency matrix using soft power and topological overlap matrix (TOM). Soft thresholding process transforms the correlation matrix to mimic the scale-free topology. TOM is used to filter weak connections during network construction. Module identification is based on TOM and on average linkage hierarchical clustering. The soft power  $\beta = 20$  ( $R^2 = 0.9160$ ) was chosen using the scale-free topology criterion (Supplementary Fig. S3). Finally, Dynamic Tree Cut algorithm<sup>6</sup> was used for dendrogram's branch selection. The module eigengene (ME) is defined as the first principal component of a given module, which can be considered a representative of the gene expression profiles in a module. Module Membership (MM), also known as eigengene-based connectivity ( $kME$ ), is defined as the correlation of each gene expression profile with the module eigengene of a given module.

**Module-trait association.** Module-trait association analysis was accomplished using the WGCNA package (version 1.69-81)<sup>16</sup> in R environment (version 3.6.2)<sup>10</sup>. For this analysis we considered as specific traits: severity group, age subgroup, sex, race, comorbidities, therapy, clinical laboratory characteristics, hospital stay, and blood collection time point after first symptoms onset. Only traits present in three or more patients were considered in the module-trait association analysis. Subsequently, the gene significance (GS), i.e., a value for the correlation between specific traits and gene expression profiles<sup>16</sup> was obtained. The mean GS value for a particular module is considered as a measure of module significance (MS). Modules presenting a significant  $p$ -value ( $p < 0.05$ ) and a positive trait correlation were selected for functional analysis.

**Node categorization.** All genes belonging to a given module were designated here as ME genes (for module eigengene gene). The modules significantly correlated to specific traits (age group, clinical, and laboratorial data) were further evaluated for identifying relevant hub genes, i.e., the highly connected genes, here termed high hierarchy (HH) genes, that hold the transcriptional network together and are also associated to specific cellular processes or link different biological processes<sup>17</sup>. Thus, connectivity measures were used for the hierarchical categorization of hub genes, considering connectivity values related to the network (overall connectivity) and to the module (intramodular connectivity for each gene based on its Pearson's correlation with all other genes in the module).

Intramodular node connectivity was calculated considering: (i)  $kTotal$ , the whole network connectivity of each gene; (ii)  $kWithin$ , gene connections with other genes in the same module<sup>16</sup>. Genes presenting high  $kTotal$  and  $kWithin$  are classified as high hubs (Hhubs), genes presenting high  $kTotal$  but low  $kWithin$  are called

eHubs, and genes presenting high  $k_{\text{Within}}$  but low  $k_{\text{Total}}$  are the iHubs. The iHubs connect most of the genes in a transcriptional module, whereas the eHubs connect different transcriptional modules, and the Hhubs hold the transcriptional modules and the network together<sup>18</sup>. The top 10 genes presenting the highest  $k_{\text{Total}}$  and/or  $k_{\text{Within}}$  values were selected as HH genes. All gene values were plotted in a  $k_{\text{Total}}$  (x-axis) vs.  $k_{\text{Within}}$  (y-axis) graph. Additionally, the expression profile was assessed for the selected hubs through GS values, i.e., the GS of the  $n$ th gene is the correlation measure between the  $n$ th gene expression and the specific trait. Positive or negative GS values mean that the  $n$ th gene is hyper- or hypo-expressed for the specific trait. Only GS values with  $p < 0.01$  were considered significant for the trait.

**Enrichment analyses.** WGCNA modular gene set enrichment analyses for Gene Ontology Biological Process (GO BP), KEGG pathways, and Reactome pathways were accomplished by using the Enrichr online web-based tool<sup>19</sup>. The terms presenting adjusted  $p$ -value  $< 0.05$  for the modules or  $p$ -value  $< 0.05$  for the hub genes, were considered significantly enriched. The same enrichment analysis strategy was applied for all hub genes and DE genes.

**Statistical analysis.** Significance analysis for microarray (SAM) using MeV software (version 4.9.0) was used for differential gene expression analyses. The relative expression of the differentially expressed genes (DEGs) was normalized with the endogenous reference gene *GUSB* for statistical analysis (Supplementary Table S3). The t-test was used for statistical analyses of differentially expressed genes. Mann Whitney, Chi-square, and Fisher exact tests were used for statistical analysis of the clinic-demographical data. All statistical analyses were performed in GraphPad Prism (version 8). The FAMD analysis was performed as described in “Factor analysis of mixed data (FAMD)” section.

**Gene co-expression network (GCN).** The GCNs were constructed by using Pearson's correlation. Data analysis and visualization were accomplished through Cytoscape<sup>20</sup> version 3.9.0.

## Results

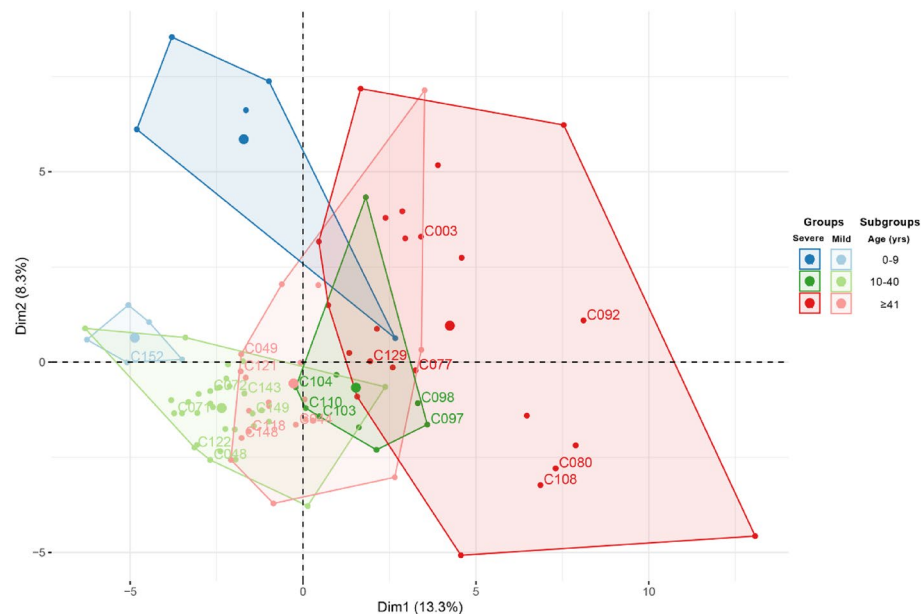
**Patient demographic and clinical characteristics.** The clinical and demographic characteristics of the severe (hospitalized) and mild (oligosymptomatic) COVID-19 patients are shown in Table 1. The median age for the Severe and Mild groups was significantly different: 50 years and 34 years, respectively. There was a predominance of males in the Severe group as compared to the Mild group. There were more young male patients (0–40 years of age) in the Severe group. Significant lower median hospital stay was observed for young patients (10–40 years of age) as compared with elderly patients (Table 1). In the Severe group more patients presented fever, dyspnea, wheezing in the chest, and vomiting, while more oligosymptomatic individuals presented coryza and headache (Table 1). A significantly higher number of patients in the Severe group presented at least one comorbidity. In decreasing order of frequency: systemic arterial hypertension (SAH), obesity, diabetes mellitus (DM), chronic heart conditions, chronic lung diseases, and chronic kidney disease (Supplementary Fig. S4).

The hemogram results were interpreted as normal, high, or low, based on reference values for age and sex (Supplementary Table S4). The comparative analysis between the Severe and Mild groups showed that more individuals in the Severe group presented high values for segmented neutrophils and low values for erythrocytes, hematocrit, hemoglobin, lymphocytes, and immature neutrophils (Supplementary Fig. S5). Severe patients had significantly elevated values for all hemogram-derived ratios when compared to Mild (i.e., oligosymptomatic) patients (Supplementary Table S5), and higher values were found for PLR and SII parameters in the subgroups B and D (patients over 41 years of age), thus confirming the positive correlation of elevated hemogram-derived ratios with inflammation and severe COVID-19<sup>10</sup>. Additionally, the individuals in the Severe group expressed higher levels of IL-6 and IL-10 when compared with the Mild group. The cytokines IL-4, IL-2, and IFN- $\gamma$ , were also elevated in these individuals (Supplementary Fig. S6).

FAMD analysis for clinical characteristics yielded six dimensional clusters for patients aged 0–9, 10–40 and over 40 years, in the Severe and Mild groups (Fig. 1). All patients selected for transcriptomics studies are contained in four dimensional clusters: 10–40 and over 40 years for both Severe and Mild groups.

**WGCNA and module-trait correlation analysis.** The normalized gene expression data of 6375 GO annotated genes were used for network construction and module identification by WGCNA. Nine transcriptional modules were identified. Module sizes ranged from 147 genes in the magenta module to 891 genes in the turquoise module (Supplementary Fig. S7). The hierarchical clustering dendrogram of module-eigengenes revealed two meta-modules. In the meta-module I, which encompasses the blue, red, turquoise, magenta, pink, black, and yellow modules, the transcriptional modules are correlated with severe phenotypes, whereas the in the meta-module II, that includes the brown and green modules, the correlation is with mild phenotypes (Supplementary Fig. S8a). The module-trait correlation analysis identified three transcriptional modules with at least one significant correlation ( $p < 0.05$ ) with disease severity (Severe; Mild) or with the A and B age subgroups (Supplementary Figs. S7, S8b). The yellow module is highly ( $r = 0.70$ ) and positively correlated with the Severe group and with high level of segmented neutrophils ( $r = 0.72$ ), and weakly ( $r = 0.46$ ) correlated with subgroup B (old and severe patients) and with low level of lymphocytes ( $r = 0.51$ ). The magenta module is positively correlated with the Severe group ( $r = 0.44$ ), subgroup A (young and severe patients,  $r = 0.47$ ), and high level of segmented neutrophils ( $r = 0.46$ ). It is also positively correlated with male sex in the Severe group ( $r = 0.47$ ) and with white race ( $r = 0.55$ ), and negatively correlated with DM in the Severe group ( $r = -0.62$ ). The black module is positively correlated with subgroup A ( $r = 0.48$ ) and with high level of segmented neutrophils ( $r = 0.48$ ), and negatively correlated with DM in the Severe group ( $r = -0.63$ ).





**Figure 1.** Exploratory multivariate analysis. Factor analysis of mixed data (FAMD) with grouping variables identified four dissimilar groups: the two main groups (Severe and Mild), and the two age-related subgroups (10–40 years old; 41–85 years old).

We also performed a functional characterization of these three transcriptional modules and identified their high hierarchy (HH) genes (Supplementary Fig. S9). Transcriptional modules often represent biological processes that can be phenotype-specific<sup>18</sup>. The functional enrichment among the genes within a module is widely used for disclosing its biological meaning<sup>18</sup>.

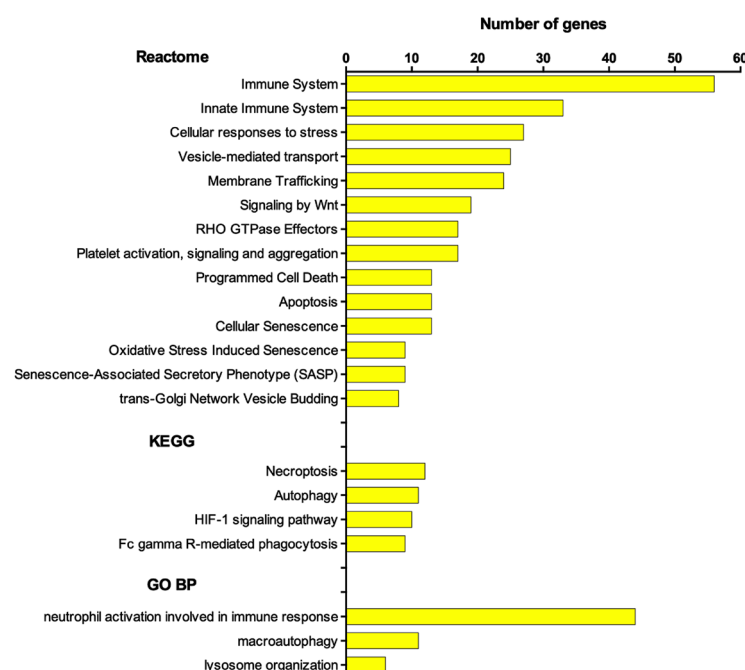
**Yellow module.** The enrichment analyses for this module (Fig. 2a, Supplementary Table S6) showed that the most represented cellular processes and functional pathways mainly reflect the inflammatory and innate immune responses against SARS-CoV-2<sup>21</sup>. Two terms are worth mentioning: (i) in the GO BP analysis, the most represented category is related to neutrophil activation (44 genes), what is quite expectable due to the role of neutrophils in restricting viral replication and diffusion<sup>22</sup>; (ii) in the KEGG analysis, the genes of the HIF-1 signaling pathway, that are hyper-expressed in the peripheral blood mononuclear cells (PBMC) of COVID-19 patients<sup>23</sup> and aggravate inflammatory responses<sup>24</sup>.

The yellow module harbors 14 HH genes (Fig. 2b). The Hhubs *TMEM59*<sup>25</sup> and *ATP6V1E1*<sup>26</sup>, and the iHubs *SPTLC1*<sup>27</sup> and *SNAPIN*<sup>28</sup> are autophagy-related genes, the latter being critical for autophagosome maturation in macrophages. Two Hhubs—*PRDX3*<sup>29</sup> and *SDHB*<sup>30</sup>—are involved in the protection against oxidative stress. The eHub *TXNDC12* (alias *ERP16*) takes part in the cellular defense against prolonged ER stress<sup>31</sup>. A decrease in ER-stress was shown to occur in a stage-specific manner during neutrophil and macrophage differentiation<sup>32</sup>. The iHub *C1GALT1C1* codes for a molecular chaperone (Cosmc) that plays a crucial role TRAIL-induced apoptosis<sup>33</sup>. A well-balanced IFN/TRAIL response is necessary for overcoming viral infections<sup>34</sup>. Finally, it is important to mention that in this module seven HH genes also have high GS values. Four of these genes—*TMEM59*, *PRDX3*, *SPTLC1*, and *C1GALT1C1*—are positively correlated with the Severe group and with high level of segmented neutrophils. Two other genes—*EIF4A3*, a modulator of the non-sense mediated mRNA decay pathway<sup>35</sup> and *SNAPIN*—are positively correlated with the Severe group. *TENT2* (alias *PAPD4*), a gene involved in miRNA processing<sup>36</sup>, is positively correlated with high level of segmented neutrophils. Taken together, the gene enrichment analyses and the functional characterization of the HH genes for the yellow module fairly agree with the positive correlation of this module with COVID-19 severe cases, with subgroup B (aged and severe patients), and with high number of segmented neutrophils and lymphopenia (i.e., elevated NLR and SII; Supplementary Table S5), two hallmarks of COVID-19 severity<sup>37</sup>.

**Magenta module.** The enrichment analyses revealed a predominance of pathways related with innate and adaptive immune responses and inflammation (Fig. 3a, Supplementary Table S7). Moreover, the Fc epsilon receptor signaling pathway indicates the involvement of basophils and, indeed, it was recently shown that these cells play an active role in the immunity against SARS-CoV-2<sup>38</sup>.

The magenta module has 13 HH genes (Fig. 3b) and most of them act on the modulation of the innate and adaptive antiviral immune responses. The Hhub *PPP2CA* codifies protein phosphatase 2A, an important regulator of inflammatory signaling<sup>39</sup>. The Hhub *RNF114* codes for a RING-type zinc-finger protein that modulates NF-κB activity and T-cell activation<sup>40</sup>. The Hhub *HTR3A* encodes the subunit A of the 5-HT3 receptor of serotonin, found in monocytes, B and T cells and probably involved in their recruitment to sites of inflammation<sup>41</sup>.

a

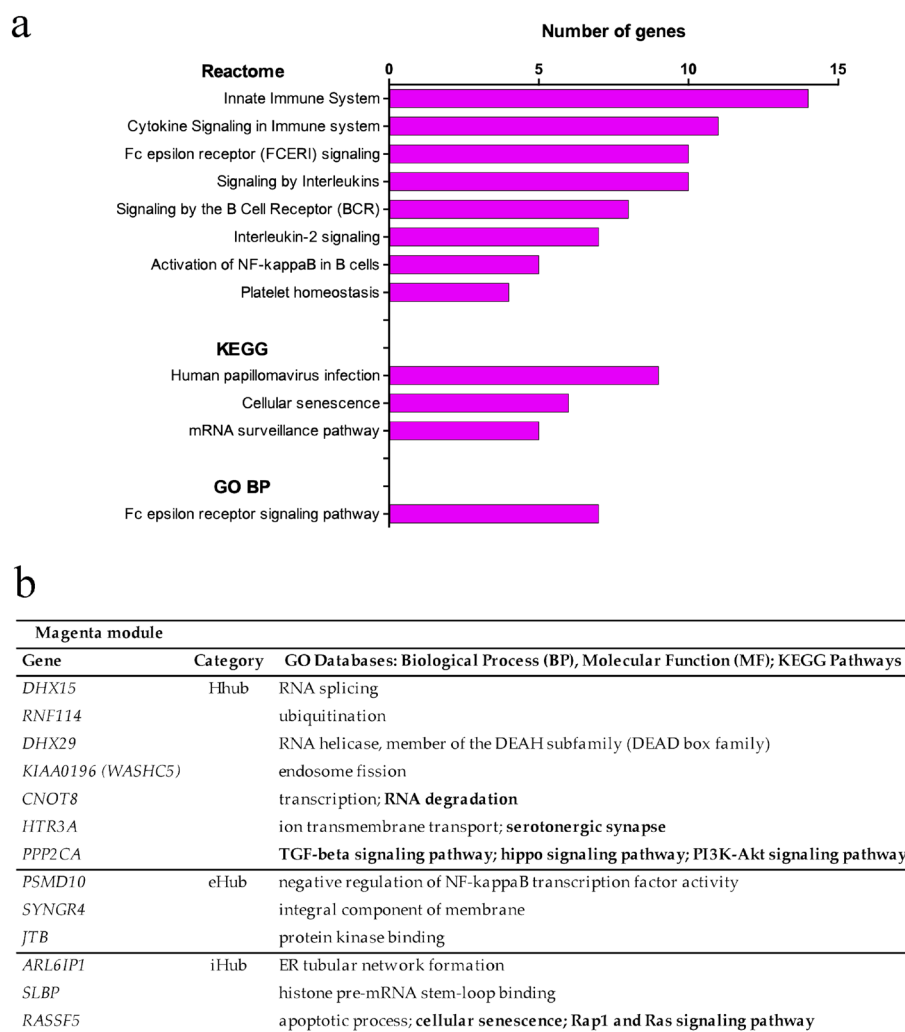


b

Yellow module			GS Values	
Gene	Category	GO Databases: Biological Process (BP), Molecular Function (MF); KEGG Pathways	Severe	H-SN
<i>TMEM59</i>	Hhub	autophagy	0.59	0.62
<i>DECR1</i>		fatty acid beta-oxidation		
<i>ATP6V1E1</i>		regulation of macroautophagy; mTOR signaling pathway		
<i>PRDX3</i>		cellular response to oxidative stress	0.59	0.59
<i>SDHB</i>		<b>oxidative phosphorylation</b>		
<i>METTL6</i>		tRNA methylation		
<i>EIF4A3</i>	eHub	mRNA splicing, via spliceosome	0.59	
<i>COPS5</i>		ubiquitination; regulation of stress-activated MAPK cascade		
<i>PAPD4 (TENT2)</i>		histone mRNA catabolic process		0.62
<i>TXNDC12</i>		endoplasmic reticulum stress-induced intrinsic apoptosis		
<i>SNAPIN</i>	iHub	autophagosome maturation	0.60	
<i>MBD2</i>		methylation-dependent chromatin silencing		
<i>SPTLC1</i>		macroautophagy; <b>sphingolipid metabolism</b>	0.66	0.70
<i>C1GALT1C1</i>		<b>mucin type O-glycan biosynthesis</b>	0.61	0.62

**Figure 2.** Enrichment analysis and high hierarchy genes (HH genes) for the yellow module. **(a)** Histogram of enriched Reactome, KEGG pathways, and GO BP terms. The terms with adjusted  $p < 0.05$  were considered significant. **(b)** HH genes (Hhubs, iHubs, and eHubs) of this module, which is significantly and positively correlated with Severe group, subgroup B, H-SN, and L-Lymphs traits. Each HH gene is identified by its hierarchical categorization, GO biological process or molecular function, and KEGG Pathways-related terms (in bold letters). Only positive (i.e., hyper-expressed) or negative (i.e., hypo-expressed) significant GS values for the specific trait ( $p < 0.01$ ) are shown. *H-SN* high percentage of segmented neutrophils in the hemogram, GS gene Significance.

Decreased serum levels of serotonin were found in severe cases of COVID-19, whereas in the mild cases the serotonin levels were close to those found in control individuals<sup>42</sup>. The Hhubs *DHX15*<sup>43</sup> and *DHX29*<sup>44</sup> code for two DEAH-box helicases that are key sensors for antiviral defense against RNA virus infection. The Hhub *WASHC5* codes for strumpellin, a protein involved in endosomal fission<sup>45</sup> that was shown to interact directly with SARS-CoV-2<sup>46</sup>. The Hhub *CNOT8* encodes the CNOT8 catalytic subunit of the CCR4-NOT complex, which is required for selective mRNA degradation (mRNA surveillance) and for preventing excessive inflammatory responses<sup>47</sup>. The eHub *PSMD10* codes for gankyrin and promotes autophagy<sup>48</sup>, a crucial innate immune response against infection. The eHub *SYNGR4* codes for the integral membrane component synaptogyrin and promotes endosome recycling<sup>49</sup> that is essential for cytokine release by neutrophils<sup>50</sup>. The eHub *ITB* (alias *PAR*) acts on the regulation of mitochondrial function<sup>51</sup> and the low expression of mtDNA in the immune system cells of COVID-19 patients was recently reported<sup>52</sup>. Lastly, there are three iHubs: *SLBP*, which encodes a TNF-induced stem-loop binding protein that regulates histone metabolism, inflammation, and viral replication<sup>53</sup>. *RASSF5*



**Figure 3.** Enrichment analysis and high hierarchy genes (HH genes) for the magenta module. **(a)** Histogram of enriched Reactome, KEGG pathways, and GO BP terms. The terms with adjusted  $p < 0.05$  were considered significant. **(b)** HH genes (Hhubs, iHubs, and eHubs) of this module, which is significantly and positively correlated with Severe group, subgroup A, and with H-SN traits. Each HH gene is identified by its hierarchical categorization, GO biological process or molecular function, and KEGG Pathways-related terms (in bold letters). This module did not present significant Gene Significance (GS) values for any specific trait ( $p < 0.01$ ).

(alias *NORE1A*), an apoptosis inducer and senescence effector<sup>54</sup>; and *ARL6IP1*, that codes for an ER-shaping protein/ involved in the fine control of ER organization<sup>55</sup>.

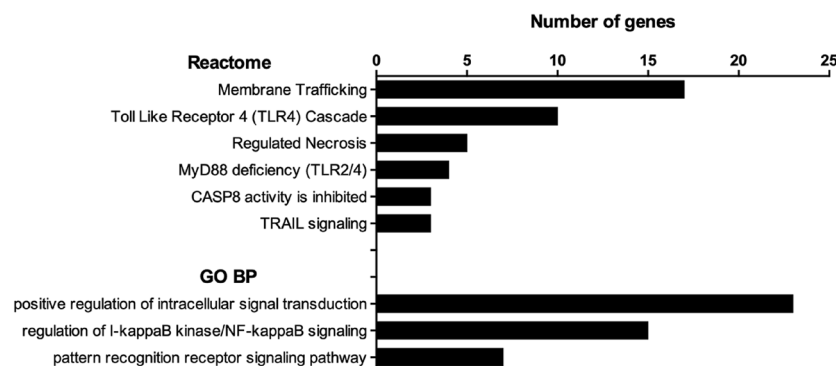
**Black module.** The enriched pathways in the black module (Fig. 4a, Supplementary Table S8) reflect some of the main mechanisms involved in the innate immune and inflammatory responses to SARS-CoV-2. Many pathways, such as TLR/4, MyD88/TLR2, CASP8, TRAIL signaling, and necrosis (Fig. 4a) are mechanistically related and take part in the recognition of viral proteins, inflammatory responses, and cell death<sup>56</sup>. Interestingly, the increased expression of TLRs and MyD88 were found to be positively correlated with COVID-19 severity<sup>57</sup>. Noteworthy, the GO BP analysis shows enrichment for I-kappaB Kinase/NF-kappaB signaling pathway, a master regulator of inflammatory and immune responses<sup>58</sup>.

This module has 13 HH genes (Fig. 4b) of which five are directly related to immune response regulation. The Hhub *CACNA1C* codes for the alpha-1 subunit of CaV1.2 calcium channel and is involved in the regulation of human Th2-lymphocyte functions<sup>59</sup>. The eHub *RAC1* codifies a Rho GTPase that triggers NF-kB activation<sup>60</sup>. The eHub *PTMA* encodes prothymosin-alpha, an alarmin involved in cell proliferation, apoptosis, and immune regulation and whose expression in CD8 T memory stem cells is increased in COVID-19<sup>61</sup>. The iHub *FOXP4* codifies a transcription factor required for T cell recall response to pathogens<sup>62</sup>. The iHub *UCP3* codes for an uncoupling protein (UC) involved in the maintenance of Th17/Treg cell balance<sup>63</sup>, which is skewed towards Th17 in COVID-19<sup>64</sup>.

The other HH genes in the black module are involved in different biological functions, some of them possibly related to modulatory mechanisms acting on inflammatory and immune responses. The Hhub *CGYB* codes



a



b

Black module			GS Value
Gene	Category	GO Databases: Biological Process (BP), Molecular Function (MF); KEGG Pathways	H-SN
<i>EVX1</i>	Hhub	transcription by RNA polymerase II; transcriptional repressor during embryogenesis	
<i>SHISA4</i>		protein binding	
<i>CYGB</i>		response to oxidative stress	
<i>LOC113230 (MIS)</i>		MISP family member 3	
<i>FHL1</i>		potassium ion transport; negative regulation of cell growth; JAK-STAT signaling	
<i>CACNA1C</i>	eHub	calcium signaling pathway	-0.58
<i>HBZ</i>		zeta-globin	
<i>PTMA</i>		negative regulation of apoptotic process; transcription	
<i>RAC1</i>		Rap1 signaling pathway; Ras signaling pathway	
<i>FOXP4</i>	iHub	transcription	
<i>DPEP3</i>		proteolysis	
<i>UCP3</i>		mitochondrial membrane transport	
<i>PYY2</i>		pseudogene	

**Figure 4.** Enrichment analysis and high hierarchy genes (HH genes) for the black module. **(a)** Histogram of enriched Reactome and GO BP terms. The terms with adjusted  $p < 0.05$  were considered significant. **(b)** HH genes (Hhubs, iHubs, and eHubs) of this module, which is significantly and positively correlated with subgroup A and with H-SN traits. Each HH gene is identified by its hierarchical categorization, GO biological process or molecular function, and KEGG Pathways-related terms (in bold letters). Only positive (i.e., hyper-expressed) or negative (i.e., hypo-expressed) significant GS values for the specific trait ( $p < 0.01$ ) are shown. H-SN high percentage of segmented neutrophils in the hemogram, GS gene significance.

for cytoglobin, a protein that protects cells against oxidative stress<sup>65</sup>. The Hhub *FHL1* encodes a four-and-a-half-LIM-domain protein involved in gene transcription regulation, cytoarchitecture, cell proliferation, and signal transduction<sup>66</sup>. Interestingly, FHL1 inhibits the vascular endothelial growth factor (VEGF) expression<sup>67</sup> and elevated serum levels of VEGF were positively correlated with disease severity in COVID-19<sup>68</sup>. The Hhub *MISP3* codes for a protein involved in spindle orientation and mitotic progression<sup>69</sup>. The Hhub *EVX1* encodes a transcriptional repressor<sup>70</sup> and the Hhub *SHISA4* codes for a transmembrane adaptor protein<sup>71</sup>. The iHub *DPEP3* encodes dipeptidase 3, involved in the hydrolytic metabolism of dipeptides, and the iHub *PYY2* is a pseudogene coding for peptide YY2. Finally, the eHub *HBZ* codes for zeta hemoglobin and HBZ expression may be interpreted as evidence for circulating erythroid precursors in hospitalized (requiring oxygen) patients<sup>72</sup>, being negatively correlated with hemoglobin levels. *HBZ* has a negative GS value for the H-SN trait (Fig. 4b).

**Differential gene expression analyses.** Two comparative gene expression analyses (SAM) were conducted: (i) Severe ( $n = 11$ ) vs. Mild ( $n = 12$ ) groups; (ii) A ( $n = 5$ ) vs. B ( $n = 6$ ) subgroups. In the Severe vs. Mild comparison 50 differentially expressed genes (DEGs) were found, of which 49 were hyper-expressed in the Severe group, with fold-changes ranging between 8.8 and 2.0. The A vs. B comparison yielded only two DEGs. Subsequently, a normalized expression analysis (using *GUSB* as endogenous reference) showed that these two genes did not differ significantly between the subgroups A and B.

The enrichment analysis for the DEG set of the Severe group showed that these genes are mainly related with neutrophil activation, innate immune response, and inflammation (Supplementary Fig. S10, Table S9).

A normalized gene expression analysis (using *GUSB* as an endogenous reference gene) was then performed for the DEGs firstly selected according to KEGG and GO BP enrichment analyses. It was found that 23 genes in the Severe group were significantly ( $p < 0.005$ ) hyper-expressed (Table 2). Noteworthy, 17 out of the 23 genes hyper-expressed in the Severe group had been previously reported to be hyper-expressed in the PBMC from COVID-19 patients according to the COVID-19 Related Gene Sets, a database in the Enrichr webtool library<sup>19</sup>.

Comparison	DEG	Gene expression		Database for enrichment analysis
		Fold-change	Relative ( <i>p</i> -value)*	GO BP or KEGG (in <i>italic</i> )
Severe × Mild	<i>LTF</i>	8.53	0.0001	Neutrophil activation involved in immune response; negative regulation of apoptotic process
	<i>HP**</i>	6.95	0.0002	Neutrophil activation involved in immune response; positive regulation of cell death
	<i>CEACAM8</i>	6.16	0.0009	Neutrophil activation involved in immune response
	<i>HPR**</i>	5.64	0.0003	Acute inflammatory response; positive regulation of cell death
	<i>LCN2</i>	5.02	0.0013	Neutrophil activation involved in immune response
	<i>ARG1</i>	4.98	0.0009	Neutrophil activation involved in immune response
	<i>GYGI**</i>	4.06	0.0002	Neutrophil activation involved in immune response
	<i>MPO</i>	4.03	0.0011	Neutrophil activation involved in immune response; negative regulation of apoptotic process
	<i>ORM1</i>	3.98	0.0008	Neutrophil activation involved in immune response; platelet degranulation
	<i>TXNDC5</i>	3.66	0.0020	Neutrophil activation involved in immune response; negative regulation of apoptotic process
	<i>ORM2**</i>	3.56	0.0006	Neutrophil activation involved in immune response; platelet degranulation
	<i>CAMP</i>	3.27	0.0006	Neutrophil activation involved in immune response
	<i>RRM2**</i>	2.95	0.0002	p53 signaling pathway
	<i>GGH**</i>	2.86	0.0004	Neutrophil activation involved in immune response
	<i>SELP</i>	2.71	< 0.0001	Platelet degranulation
	<i>CCNB1**</i>	2.57	0.0004	<b>p53 signaling pathway; cellular senescence</b>
	<i>BMP6**</i>	2.43	0.0001	<b>Hippo signaling pathway</b>
	<i>BIRC5**</i>	2.37	< 0.0001	<b>Hippo signaling pathway; negative regulation of apoptotic process</b>
	<i>SNCA**</i>	2.33	0.0004	Positive regulation of cell death; negative regulation of oxidoreductase activity
	<i>YWHAH**</i>	2.25	0.0005	<b>Hippo signaling pathway</b>
	<i>CCNB2**</i>	2.23	< 0.0001	<b>p53 signaling pathway; cellular senescence</b>
	<i>STOM**</i>	2.22	0.0004	Neutrophil activation involved in immune response
	<i>CDK1</i>	2.14	< 0.0001	<b>p53 signaling pathway; cellular senescence; negative regulation of apoptotic process</b>

**Table 2.** DEGs selected as potential biomarkers for the Severe group and for the subgroup C. Genes in bold are hyper-expressed in the PBMC from COVID-19 patients (COVID-19 Related Gene Sets/Enrichr database). \*t-test for relative expression of the DEGs normalized with endogenous reference gene *GUSB*. \*\*Newly identified as a potential COVID-19 transcriptional biomarker.

(Table 2). It is also important to mention that 13 out of the 23 genes hyper-expressed in the Severe group are related with neutrophil activation.

**DEGs as transcriptional biomarkers.** The significant fold-change values and the biological functions of all DEGs found for the Severe group (Table 2, Supplementary Figs. S11, S12) clearly show their potentiality as transcriptional biomarkers. The ascribed roles of these genes—of which 13 are newly identified biomarkers (Table 2)—in the immune response to COVID-19 are addressed in the following paragraphs.

Among the seven genes in the GO BP category “neutrophil activation involved in immune response” (Supplementary Fig. S11), three were already identified as COVID-19 biomarkers: *CEACAM8* (alias *CD66B*), that codes for a neutrophil cell-adhesion protein and is highly expressed in patients with severe COVID-19<sup>73</sup>; *ARG1*, coding for arginase 1 and up-regulated in severe cases of COVID-19<sup>74</sup>; and *LCN2*, which encodes lipocalin 2, a marker of neutrophil activation that was classified by machine learning algorithm as one the most potent discriminators of critical illness in COVID-19<sup>75</sup>. One DEG in this category, *CAMP*, codify for the antimicrobial molecule cathelicidin (LL37), a modulator of TLR activation and inflammation that has been proposed as a potential candidate for COVID-19 prevention and treatment<sup>76</sup>. The remaining three DEGs are related to neutrophil immune functions and metabolism: *STOM*, that codes for stomatin, a membrane protein associated with azurophilic granules<sup>77</sup>; *GGH*, that encodes gamma-glutamyl hydrolase, a lysosomal enzyme involved in the immune response of neutrophils<sup>78</sup>; and *GYGI* that codifies glycogenin, an enzyme involved in the synthesis of glycogen<sup>79</sup>.

In the GO BP category “neutrophil activation and positive regulation of cell death” (Supplementary Fig. S6) there are two DEGs. One is *HP* that codes for haptoglobin, a hemoglobin-binding plasma protein stored and released by neutrophils in response to activation<sup>80</sup>. Haptoglobin plasmatic levels were shown to correlate with disease severity in COVID-19, being reduced in critical patients<sup>81</sup> and elevated in COVID-19 children<sup>82</sup>. The other DEG, *HPR*, codes for a haptoglobin-related protein that binds hemoglobin was shown to be elevated in the sera of children with bacterial and viral pneumonia<sup>83</sup>. In the GO category “positive regulation of cell death and negative regulation of oxidoreductase activity” there is only one DEG, *SCNA*, which codes for alpha-synuclein,

a neuropeptide expressed in brain and also in mononuclear blood cells<sup>84,85</sup> that was shown to exhibit potent antiviral activity and capacity for signaling the immune system by attracting neutrophils and macrophages and activating dendritic cells<sup>86,87</sup>.

There are three DEGs in the GO BP category “neutrophil activation and negative regulation of apoptosis” (Supplementary Fig. S11). One is *LTF*, which codifies for lactoferrin, a relevant player in innate immunity with antiviral effects against SARS-CoV-2 and a wide range of viral species<sup>88</sup>. Serum levels of lactoferrin were found to be elevated in severe cases of COVID-19<sup>89</sup>. The other DEG is *MPO*, a gene that codes for myeloperoxidase, a leukocyte-derived enzyme whose plasmatic levels are elevated in mild to severe cases of COVID-19 and down-regulated in patients with very severe disease<sup>90</sup>. The third gene is *TXND5*, that codes for the thioredoxin domain containing 5 (TXNDC5), an endoplasmic reticulum-resident protein that belongs to the thioredoxin family. The plasmatic levels of this protein are markedly elevated in septic patients, and it has been considered a therapeutic target for attenuating inflammatory responses<sup>91</sup>. Additionally, it was shown that *TXND5* is hyper-expressed in B cells after COVID-19 vaccination, being a marker for seroconversion<sup>92</sup>. Therefore, the three genes are potential candidates for monitoring disease severity in COVID-19.

The GO BP category “platelet degranulation and neutrophil activation” has three DEGs (Supplementary Fig. S11). One of these genes is *SELP*, that codes for P-selectin, a platelet cell-adhesion molecule. Increased levels of P-selectin were found in severe cases of COVID-19, what contributes for a prothrombotic state in these patients<sup>93</sup>. The other two genes in this category are *ORM1* and *ORM2*, and both are involved in the encoding of human orosomucoid protein, a major acute-phase plasma protein<sup>94</sup>. Recently, a proteomic analysis of serum from COVID-19 patients showed a significant down-regulation of the *ORM1* protein<sup>95</sup>.

The KEGG pathway “p53 signaling and cellular senescence” (Supplementary Fig. S12) includes four DEGs. One of these genes is *CDK1* that codes for the cyclin-dependent kinase 1 and is a master regulator of autophagy<sup>96</sup>. *CDK1* was shown to be, through bioinformatics and machine learning approaches, a relevant Hub gene in the WBC transcriptome of COVID-19 patients, with high biomarker and therapeutic target potentials<sup>97</sup>. Other two DEGs in this pathway, *CCNB1* and *CCNB2*, coding respectively for cyclin B1 and cyclin B2, are interactors of *CDK1* and regulate the mammalian cell cycle<sup>98</sup>. The fourth DEG in this pathway, *RRM2*, interacts with *CDK1* and *CNBB1* in the p53 pathway<sup>99</sup> and was recently identified, through bioinformatic analysis, as a key gene in the infection of human intestines by SARS-CoV-2<sup>100</sup>.

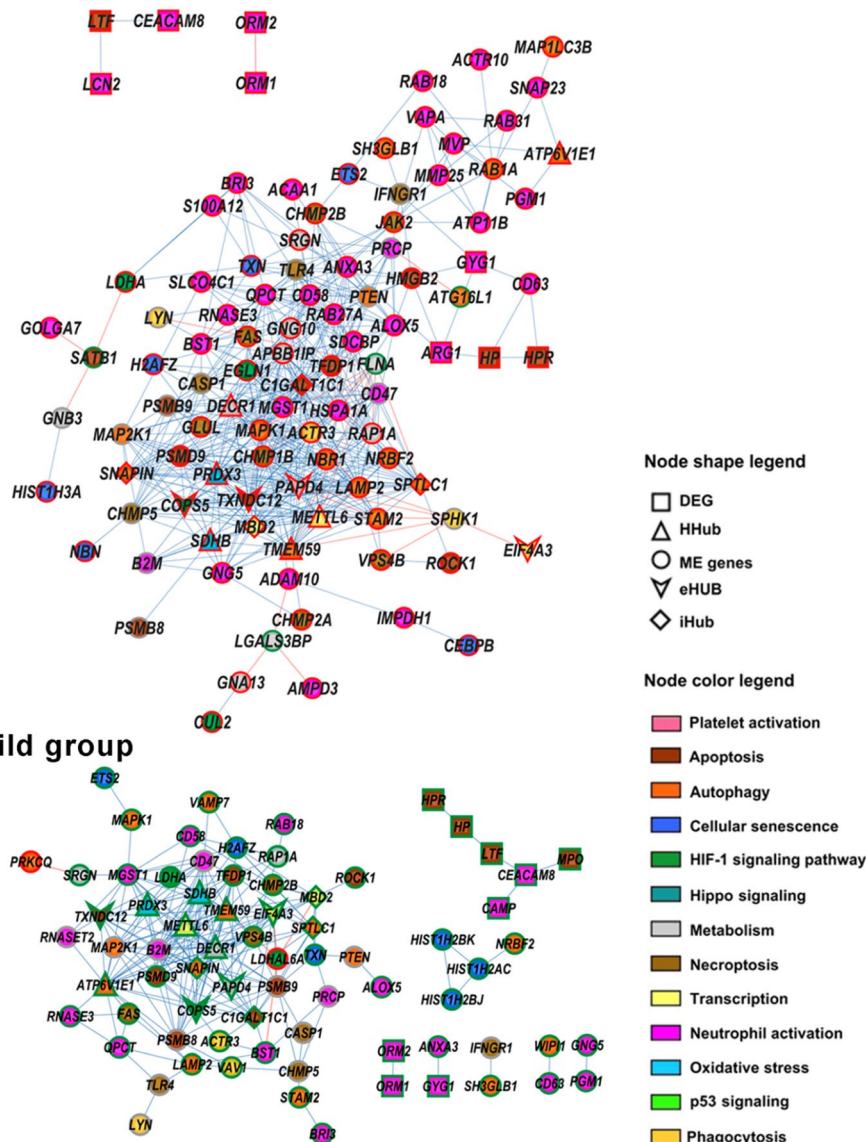
Finally, there are three DEGs in the KEGG “Hippo signaling pathway” (Supplementary Fig. S12). One of these genes is *BIRC5* that codes for survivin, an inhibitor of apoptosis protein. Survivin is indispensable for the homeostasis of the immune system, being required for innate and adaptive immune responses, differentiation of CD4+ and CD8+ memory T-cells, and for B cell maturation<sup>101</sup>. The other two DEGs are *BMP6* and *YWHAH*. The former codes for the bone morphogenetic protein 6, a regulator of vascular homeostasis and angiogenesis<sup>102</sup> recently identified as an anti-inflammatory cytokine<sup>103</sup>. The latter codes for the 14-3-3 $\eta$  (eta) protein, a phosphoserine/phosphothreonine binding protein that interacts with a wide range of protein targets and participates in multiple cellular biological functions. The 14-3-3 $\eta$  protein is involved in the modulation of antiviral defenses via the RLR signaling pathway. The interaction between 14-3-3 $\eta$  and the melanoma differentiation-associated gene 5 (*MDA5*) accelerates the activation of the *MDA5* signaling, thus helping host cells to mount an effective response against RNA viral infections<sup>104</sup>.

**Subnetworks for module genes and DEGs.** Subnetworks were constructed for the yellow and black modules. The subnetwork for the yellow module, positively correlated with the Severe group, was constructed using genes enriched for innate immune response and inflammation processes and pathways. The subnetwork for the black module, positively correlated with subgroup A, was constructed using genes enriched for inflammation, innate immune response, and immune response regulation. All these genes and their respective functions are listed in Supplementary Table S10. The DEGs identified in the Severe vs. Mild group comparison were also included in the yellow subnetwork (Supplementary Table S9). Consequently, these subnetworks were constructed with 141 genes for the yellow subnetwork and 33 genes for the black subnetwork. For all subnetworks a gene–gene link cut-off of  $r > |0.9|$  was adopted. The yellow subnetwork (Fig. 5) for the Severe group showed a higher value of connectivity (10.6) when compared with the subnetwork obtained for the Mild group (3.2). The subnetworks for the black module (Fig. 6) showed a higher connectivity for subgroup A (11.3) when compared with the subgroup B (6.0). Additionally, the yellow subnetwork contains more interconnected genes in the Severe group (104 genes and 551 links) than in the Mild group (74 genes and 241 links). It is worth to mention that the yellow subnetwork for the Severe group encompasses several genes involved in neutrophil activation (Fig. 5). Similarly, the black subnetwork for subgroup A contains many genes involved in the I-kappaB Kinase/NF-kappaB signaling, an important modulator of inflammatory and immune responses (Fig. 6).

## Discussion

The transcriptomic response of human leukocytes to SARS-CoV-2 infection was investigated by focusing the differences between mild and severe cases and between age subgroups (younger and older adults). The gain of this investigation strategy consists in combining host response transcriptomic data with clinical, demographic, and laboratory information, in a systems biology approach, for studying the immune response between mild and severe COVID-19<sup>1,2,105</sup>. Here, it was possible to functionally characterize the transcriptional modules correlated with age and disease severity, as well as to identify severity-associated DEGs, as discussed below. Still, this study faced two limitations, i.e., the relatively small number of patients included in the transcriptomic study (Table S2), thus impeding a statistical evaluation of the recovery parameter in these patients, and the reduced number of infants presenting severe symptoms, preventing us to investigate the leukocyte transcriptional response to SARS-CoV-2 in that young population.

## Severe group

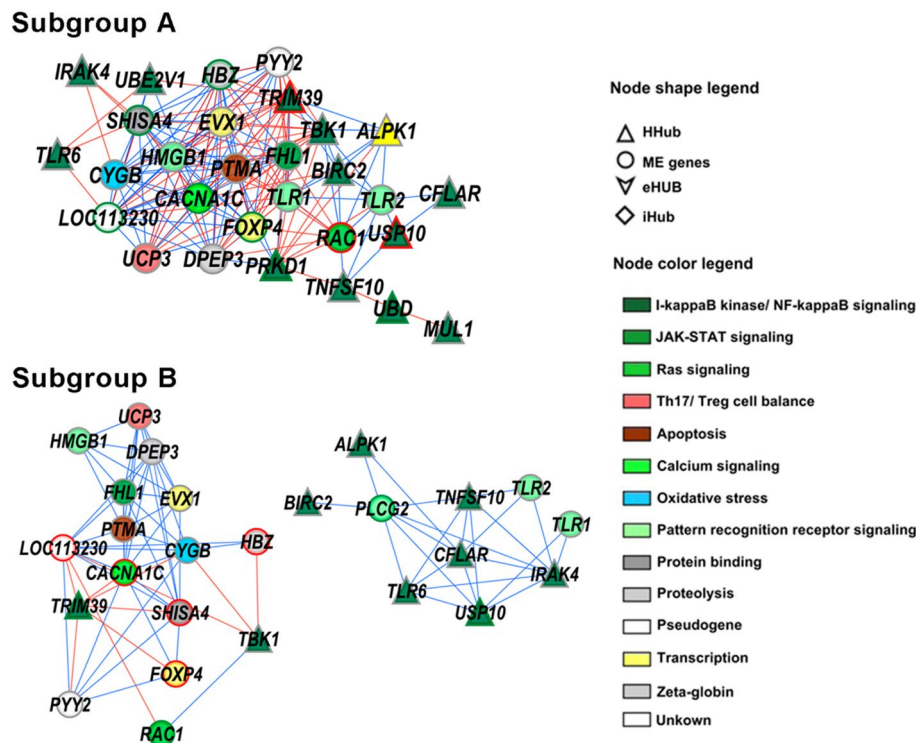


**Figure 5.** Subnetworks for the yellow module containing ME genes (non-HH module genes), hub genes, and DEGs for the Severe and Mild groups. Red or blue links indicate positive or negative expression correlation values, respectively. Red or green border colors account for hyper-expressed or hypo-expressed genes, respectively.

Initially, it is worth to recall that most of the patients in the Severe group presented fever, dyspnea, wheezing in the chest, chills, and vomiting, while in the Mild group a sizable number of patients solely presented mild symptoms, such as coryza and headache. It is well known that when the response to respiratory viruses is inadequate the infection spreads to lower respiratory tract (LRT)<sup>106</sup>. An early and effective immune response in patients infected with SARS-CoV-2 can limit and eliminate the infection yet in the upper respiratory tract (URT)<sup>107</sup>. Genomic analyses shown that there is no tissue specific genetic adaptation of SARS-CoV-2 to the URT or LRT<sup>108</sup>.

The transcriptional module differences between the Severe and Mild groups reflect and are coherent with the different presentation of symptoms and with COVID-19 severity. The functional enrichment analysis of the yellow module, associated with the Severe group, revealed that this module is related with inflammation and innate immune response, such as neutrophil activation, apoptosis, necroptosis, and HIF-1 signaling pathway. Expectedly, this module contains seven out of its 14 HH genes highly correlated with disease severity and/or high level of segmented neutrophils. Increased neutrophil count, immature circulating neutrophils, and neutrophil activation transcriptomic signatures are typically found in severe cases of COVID-19<sup>37</sup>. Interestingly, the HIF-1 pathway is associated with neutrophil activation, prolonged lifetime, and excessive function in COVID-19 severe patients<sup>24,37</sup>.





**Figure 6.** Subnetworks for the black module containing ME genes (non-HH module genes), hubs, and DEGs for the subgroups A and B. Red or blue links indicate positive or negative expression correlation values, respectively. Red or green border colors account for hyper-expressed or hypo-expressed genes, respectively.

Furthermore, the subnetwork constructed for the yellow module showed that most of the highly interconnected genes in the Severe group are hyper-expressed, and that many of these genes are involved in neutrophil activation, with other genes involved in necroptosis, apoptosis, and autophagy. This high network connectivity indicates that in the Severe group the biological functions and/or cellular processes related to innate immune response and inflammation are more activated than in the Mild group, shedding some light on the underlying transcriptomic mechanism, since the immune response to SARS-CoV-2 is characterized by neutrophil hyperactivation and high neutrophil counts<sup>109</sup>. Most of the patients of the Severe group included in this study presented high NLR as compared with the Mild group.

It is important to highlight that the young (subgroup A) and old patients (subgroup B) in the Severe group also presented differences in their transcriptomic profiles. Two modules—magenta and black—that are correlated with the subgroup A, contains many HH genes related to the innate immune response to viral infections, as well as other HH genes related to modulatory mechanisms acting on these responses. Moreover, in the black subnetwork the highest gene interconnectivity was found for subgroup A, thus indicating that immune modulatory functions are more activated in this subgroup when compared to subgroup B. These transcriptomic mechanisms could be related to the lower median hospital stay that has been observed for young patients (including our cohort) when compared with elderly patients<sup>110</sup>. In the elderly patients a less effective immune modulation of COVID-19, allied to immune senescence and, eventually, to comorbidities, would lead to a worse outcome<sup>111</sup>.

A total of 23 DEGs were found, all hyper-expressed in the Severe group. These genes are potential biomarkers for COVID-19 severity (13 are newly described). Remarkably, thirteen of these DEGs are involved in neutrophil activation, confirming the role of neutrophils in severe COVID-19<sup>37</sup>. Four DEGs are involved in the p53 signaling pathway, that has been associated with lymphopenia in severe COVID-19 patients<sup>3</sup>, and three with the Hippo pathway, recently associated with the antiviral host response in COVID-19<sup>112</sup>. Epidemiological studies show that age remains a relevant predictive factor for COVID-19 severity<sup>111</sup>, together with pre-existing conditions, such as obesity, DM and hypertension<sup>113</sup>. However, the molecular mechanisms determining COVID-19 severity are not yet well understood. Hence, there is a demand for biomarkers derived from comparative transcriptome analyses of mild and severe cases, combined with patients' clinico-demographic and laboratory data. These biomarkers may well be useful for the better stratification of risk factors in COVID-19.

### Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. All microarray raw data have been deposited in GEO public database (<http://www.ncbi.nlm.nih.gov/geo>, token ahwlmiservcnfid), a MIAME compliant database, under accession number GSE193022. All data will be released with paper publication.



Received: 21 October 2022; Accepted: 16 January 2023

Published online: 17 January 2023

## References

1. Arunachalam, P. S. *et al.* Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210–1220. <https://doi.org/10.1126/science.abc6261> (2020).
2. Jackson, H. *et al.* Characterisation of the blood RNA host response underpinning severity in COVID-19 patients. *Sci. Rep.* **12**, 12216. <https://doi.org/10.1038/s41598-022-15547-2> (2022).
3. Xiong, Y. *et al.* Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg. Microbes Infect.* **9**, 761–770. <https://doi.org/10.1080/22221751.2020.1747363> (2020).
4. Daamen, A. R. *et al.* Comprehensive transcriptomic analysis of COVID-19 blood, lung, and airway. *Sci. Rep.* **11**, 7052. <https://doi.org/10.1038/s41598-021-86002-x> (2021).
5. Prokop, J. W. *et al.* High-density blood transcriptomics reveals precision immune signatures of SARS-CoV-2 infection in hospitalized individuals. *Front. Immunol.* **12**, 694243. <https://doi.org/10.3389/fimmu.2021.694243> (2021).
6. Langfelder, P., Zhang, B. & Horvath, S. *Dynamic Tree Cut: In-Depth Description, Tests and Applications*. <http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting> (2007).
7. Naveca, F. G. *et al.* COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nat. Med.* **27**, 1230–1238. <https://doi.org/10.1038/s41591-021-01378-7> (2021).
8. Pagès, J. Analyse factorielle de données mixtes. *Rev. Stat. Appl.* **52**, 19 (2004).
9. Lê, S., Josse, J. & Huisson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 18. <https://doi.org/10.18637/jss.v025.i01> (2008).
10. Team, R. C. R: A Language and Environment for Statistical Computing. <http://www.R-project.org/> (2021).
11. Velazquez, S. *et al.* Hemogram-derived ratios as prognostic markers of ICU admission in COVID-19. *BMC Emerg. Med.* **21**, 89. <https://doi.org/10.1186/s12873-021-00480-w> (2021).
12. Kleckner, I. R. *et al.* Effects of exercise on inflammation in patients receiving chemotherapy: A nationwide NCORP randomized clinical trial. *Support Care Cancer* **27**, 4615–4625. <https://doi.org/10.1007/s00520-019-04772-7> (2019).
13. Corman, V. M. *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25**, 2000045. <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045> (2020).
14. Sakthivel, S. K. *et al.* Comparison of fast-track diagnostics respiratory pathogens multiplex real-time RT-PCR assay with in-house singleplex assays for comprehensive detection of human respiratory viruses. *J. Virol. Methods* **185**, 259–266. <https://doi.org/10.1016/j.jviromet.2012.07.010> (2012).
15. Gentleman, V. *et al.* (eds) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 (Springer, 2005).
16. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
17. Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68. <https://doi.org/10.1038/nrg2918> (2011).
18. van Dam, S., Vösa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform.* **19**, 575–592. <https://doi.org/10.1093/bib/bbw139> (2018).
19. Kuleshov, M. V. *et al.* Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97. <https://doi.org/10.1093/nar/gkw377> (2016).
20. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
21. Lotfi, R., Kalmarzi, R. N. & Roghani, S. A. A review on the immune responses against novel emerging coronavirus (SARS-CoV-2). *Immunol. Res.* **69**, 213–224. <https://doi.org/10.1007/s12026-021-09198-0> (2021).
22. Ma, Y., Zhang, Y. & Zhu, L. Role of neutrophils in acute viral infection. *Immun. Inflamm. Dis.* <https://doi.org/10.1002/iid3.500> (2021).
23. Zhang, Y. *et al.* Identification of monocytes associated with severe COVID-19 in the PBMCs of severely infected patients through single-cell transcriptome sequencing. *Engineering (Beijing)*. <https://doi.org/10.1016/j.eng.2021.05.009> (2021).
24. Tian, M. *et al.* HIF-1 $\alpha$  promotes SARS-CoV-2 infection and aggravates inflammatory responses to COVID-19. *Signal Transduct. Target Ther.* **6**, 308. <https://doi.org/10.1038/s41392-021-00726-w> (2021).
25. Liu, Z. *et al.* TMEM59 interacts with TREM2 and modulates TREM2-dependent microglial activities. *Cell Death Dis.* **11**, 678. <https://doi.org/10.1038/s41419-020-02874-3> (2020).
26. Guo, H. *et al.* Atg5 disassociates the V1V0-ATPase to promote exosome production and tumor metastasis independent of canonical macroautophagy. *Dev. Cell* **43**, 716–730. <https://doi.org/10.1016/j.devcel.2017.11.018> (2017).
27. Lauterbach, M. A. *et al.* 1-Deoxysphingolipids cause autophagosome and lysosome accumulation and trigger NLRP3 inflammasome activation. *Autophagy* **17**, 1947–1961. <https://doi.org/10.1080/15548627.2020.1804677> (2021).
28. Shi, B. *et al.* SNAPIN is critical for lysosomal acidification and autophagosome maturation in macrophages. *Autophagy* **13**, 285–301. <https://doi.org/10.1080/15548627.2016.1261238> (2017).
29. Lee, Y. J. Knockout mouse models for peroxiredoxins. *Antioxidants* **9**, 020182. <https://doi.org/10.3390/antiox9020182> (2020).
30. Lemarie, A., Huc, L., Pazarentzos, E., Mahul-Mellier, A. L. & Grimm, S. Specific disintegration of complex II succinate: ubiquinone oxidoreductase links pH changes to oxidative stress for apoptosis induction. *Cell Death Differ.* **18**, 338–349. <https://doi.org/10.1038/cdd.2010.93> (2011).
31. Jeong, W., Lee, D. Y., Park, S. & Rhee, S. G. ERp16, an endoplasmic reticulum-resident thiol-disulfide oxidoreductase: Biochemical properties and role in apoptosis induced by endoplasmic reticulum stress. *J. Biol. Chem.* **283**, 25557–25566. <https://doi.org/10.1074/jbc.M803804200> (2008).
32. Tanimura, A. *et al.* Mitochondrial activity and unfolded protein response are required for neutrophil differentiation. *Cell Physiol. Biochem.* **47**, 1936–1950. <https://doi.org/10.1159/000491464> (2018).
33. Jiang, Y. *et al.* O-glycans on death receptors in cells modulate their sensitivity to TRAIL-induced apoptosis through affecting on their stability and oligomerization. *FASEB J.* **34**, 11786–11801. <https://doi.org/10.1096/fj.201900053RR> (2020).
34. Peteranderl, C. & Herold, S. The impact of the interferon/TNF-related apoptosis-inducing ligand signaling axis on disease progression in respiratory viral infection and beyond. *Front. Immunol.* **8**, 313. <https://doi.org/10.3389/fimmu.2017.00313> (2017).
35. Chen, T. C. *et al.* Host-derived circular RNAs display proviral activities in Hepatitis C virus-infected cells. *PLoS Pathog.* **16**, e1008346. <https://doi.org/10.1371/journal.ppat.1008346> (2020).
36. Snoek, B. C., Babion, I., Koppers-Lalic, D., Pegtel, D. M. & Steenbergen, R. D. Altered microRNA processing proteins in HPV-induced cancers. *Curr. Opin. Virol.* **39**, 23–32. <https://doi.org/10.1016/j.coviro.2019.07.002> (2019).
37. Reusch, N. *et al.* Neutrophils in COVID-19. *Front. Immunol.* **12**, 652470. <https://doi.org/10.3389/fimmu.2021.652470> (2021).
38. Murdaca, G. *et al.* Basophils and mast cells in COVID-19 pathogenesis. *Cells* **10**, 102754. <https://doi.org/10.3390/cells10102754> (2021).

39. Clark, A. R. & Ohlmeyer, M. Protein phosphatase 2A as a therapeutic target in inflammation and neurodegeneration. *Pharmacol. Ther.* **201**, 181–201. <https://doi.org/10.1016/j.pharmthera.2019.05.016> (2019).
40. Rodriguez, M. S. *et al.* The RING ubiquitin E3 RNF114 interacts with A20 and modulates NF- $\kappa$ B activity and T-cell activation. *Cell Death Dis.* **5**, e1399. <https://doi.org/10.1038/cddis.2014.366> (2014).
41. Herr, N., Bode, C. & Duerschmied, D. The effects of serotonin in immune cells. *Front. Cardiovasc. Med.* **4**, 48. <https://doi.org/10.3389/fcvm.2017.00048> (2017).
42. Soria-Castro, R. *et al.* Severe COVID-19 is marked by dysregulated serum levels of carboxypeptidase A3 and serotonin. *J. Leukoc. Biol.* **110**, 425–431. <https://doi.org/10.1002/JLB.4HI0221-087R> (2021).
43. Pattabhi, S., Knoll, M. L., Gale, M. & Loo, Y. M. DHX15 is a coreceptor for RLR signaling that promotes antiviral defense against RNA virus infection. *J. Interferon Cytokine Res.* **39**, 331–346. <https://doi.org/10.1089/jir.2018.0163> (2019).
44. Zhu, Q. *et al.* DHX29 functions as an RNA co-sensor for MDA5-mediated EMCV-specific antiviral immunity. *PLoS Pathog.* **14**, e1006886. <https://doi.org/10.1371/journal.ppat.1006886> (2018).
45. Derivery, E. & Gautreau, A. Evolutionary conservation of the WASH complex, an actin polymerization machine involved in endosomal fission. *Commun. Integr. Biol.* **3**, 227–230. <https://doi.org/10.4161/cib.3.3.11185> (2010).
46. Bouhaddou, M. *et al.* The global phosphorylation landscape of SARS-CoV-2 infection. *Cell* **182**, 685–712. <https://doi.org/10.1016/j.cell.2020.06.034> (2020).
47. Akiyama, T., Suzuki, T. & Yamamoto, T. RNA decay machinery safeguards immune cell development and immunological responses. *Trends Immunol.* **42**, 447–460. <https://doi.org/10.1016/j.it.2021.03.008> (2021).
48. Li, J. *et al.* Genetically incorporated crosslinkers reveal NleE attenuates host autophagy dependent on PSMD10. *Elife* **10**, 69047. <https://doi.org/10.7554/eLife.69047> (2021).
49. Marques, R. F. & Duncan, K. E. SYNGR4 and PLEKHB1 deregulation in motor neurons of amyotrophic lateral sclerosis models: Potential contributions to pathobiology. *Neural Regen. Res.* **17**, 266–270. <https://doi.org/10.4103/1673-5374.317960> (2022).
50. Srivastava, N. & Lacy, P. Trafficking of TNF via recycling endosomes in neutrophils. *All Asth. Clin. Immun.* **10**, 47. <https://doi.org/10.1186/1710-1492-10-S2-A47> (2014).
51. Kanome, T. *et al.* Characterization of Jumping translocation breakpoint (JTB) gene product isolated as a TGF- $\beta$ 1-inducible clone involved in regulation of mitochondrial function, cell growth and cell death. *Oncogene* **26**, 5991–6001. <https://doi.org/10.1038/sj.onc.1210423> (2007).
52. Medini, H., Zirman, A. & Mishmar, D. Immune system cells from COVID-19 patients display compromised mitochondrial-nuclear expression co-regulation and rewiring toward glycolysis. *iScience* **24**, 103471. <https://doi.org/10.1016/j.isci.2021.103471> (2021).
53. Li, M. *et al.* Stem-loop binding protein is a multifaceted cellular regulator of HIV-1 replication. *J. Clin. Investig.* **126**, 3117–3129. <https://doi.org/10.1172/JCI82360> (2016).
54. Donninger, H., Schmidt, M. L., Mezzanotte, J., Barnoud, T. & Clark, G. J. Ras signaling through RASSF proteins. *Semin. Cell Dev. Biol.* **58**, 86–95. <https://doi.org/10.1016/j.semcdb.2016.06.007> (2016).
55. Dong, R. *et al.* The inositol 5-phosphatase INPP5K participates in the fine control of ER organization. *J. Cell Biol.* **217**, 3577–3592. <https://doi.org/10.1083/jcb.201802125> (2018).
56. Zhang, Y. *et al.* An update on innate immune responses during SARS-CoV-2 infection. *Viruses* **13**, 102060. <https://doi.org/10.3390/v13102060> (2021).
57. Zheng, M. *et al.* TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nat. Immunol.* **22**, 829–838. <https://doi.org/10.1038/s41590-021-00937-x> (2021).
58. Solt, L. A. & May, M. J. The IkappaB kinase complex: Master regulator of NF-kappaB signaling. *Immunol. Res.* **42**, 3–18. <https://doi.org/10.1007/s12026-008-8025-1> (2008).
59. Robert, V. *et al.* Protein kinase C-dependent activation of CaV1.2 channels selectively controls human TH2-lymphocyte functions. *J. Allergy Clin. Immunol.* **133**, 1175–1183. <https://doi.org/10.1016/j.jaci.2013.10.038> (2014).
60. Tong, L. & Tergaonkar, V. Rho protein GTPases and their interactions with NFkB: Crossroads of inflammation and matrix biology. *Biosci. Rep.* **34**, 20140021. <https://doi.org/10.1042/BSR20140021> (2014).
61. Yu, K. *et al.* Dysregulated adaptive immune response contributes to severe COVID-19. *Cell Res.* **30**, 814–816. <https://doi.org/10.1038/s41422-020-0391-9> (2020).
62. Wiehagen, K. R. *et al.* Foxp4 is dispensable for T cell development, but required for robust recall responses. *PLoS ONE* **7**, e42273. <https://doi.org/10.1371/journal.pone.0042273> (2012).
63. O'Connor, E. B. *et al.* UCP3 reciprocally controls CD4<sup>+</sup> Th17 and Treg cell differentiation. *PLoS ONE* **15**, e0239713. <https://doi.org/10.1371/journal.pone.0239713> (2020).
64. De Biasi, S. *et al.* Marked T cell activation, senescence, exhaustion and skewing towards TH17 in patients with COVID-19 pneumonia. *Nat. Commun.* **11**, 3434. <https://doi.org/10.1038/s41467-020-17292-4> (2020).
65. Ou, L. *et al.* Recombinant human cytoglobin prevents atherosclerosis by regulating lipid metabolism and oxidative stress. *J. Cardiovasc. Pharmacol. Ther.* **23**, 162–173. <https://doi.org/10.1177/1074248417724870> (2018).
66. Wei, X. & Zhang, H. Four and a half LIM domains protein 1 can be as a double-edged sword in cancer progression. *Cancer Biol. Med.* **17**, 270–281. <https://doi.org/10.20892/j.issn.2095-3941.2019.0420> (2020).
67. Zhou, Z. *et al.* FHL1 and Smad4 synergistically inhibit vascular endothelial growth factor expression. *Mol. Med. Rep.* **7**, 649–653. <https://doi.org/10.3892/mmr.2012.1202> (2013).
68. Kong, Y. *et al.* VEGF-D: A novel biomarker for detection of COVID-19 progression. *Crit. Care* **24**, 373. <https://doi.org/10.1186/s13054-020-03079-y> (2020).
69. Zhu, M. *et al.* MISP is a novel Plk1 substrate required for proper spindle orientation and mitotic progression. *J. Cell Biol.* **200**, 773–787. <https://doi.org/10.1083/jcb.201207050> (2013).
70. Briata, P. *et al.* Transcriptional repression by the human homeobox protein EVX1 in transfected mammalian cells. *J. Biol. Chem.* **270**, 27695–27701. <https://doi.org/10.1074/jbc.270.46.27695> (1995).
71. Pei, J. & Grishin, N. V. Unexpected diversity in Shisa-like proteins suggests the importance of their roles as transmembrane adaptors. *Cell Signal* **24**, 758–769. <https://doi.org/10.1016/j.cellsig.2011.11.011> (2012).
72. Albitar, M., Peschle, C. & Liebhaber, S. A. Theta, zeta, and epsilon globin messenger RNAs are expressed in adults. *Blood* **74**, 629–637. <https://doi.org/10.1182/blood.V74.2.629.629> (1989).
73. Mazzitelli, I. *et al.* Immunoglobulin G immune complexes may contribute to neutrophil activation in the course of severe coronavirus disease 2019. *J. Infect. Dis.* **224**, 575–585. <https://doi.org/10.1093/infdis/jiab174> (2021).
74. Dean, M. J. *et al.* Severe COVID-19 is characterized by an impaired type I Interferon response and elevated levels of arginase producing granulocytic myeloid derived suppressor cells. *Front. Immunol.* **12**, 695972. <https://doi.org/10.3389/fimmu.2021.695972> (2021).
75. Meizlish, M. L. *et al.* A neutrophil activation signature predicts critical illness and mortality in COVID-19. *Blood Adv.* **5**, 1164–1177. <https://doi.org/10.1182/bloodadvances.2020003568> (2021).
76. Wang, C. *et al.* Human cathelicidin inhibits SARS-CoV-2 infection: Killing two birds with one stone. *ACS Infect. Dis.* **7**, 1545–1554. <https://doi.org/10.1021/acsinfectdis.1c00096> (2021).
77. Lawrence, S. M., Corriden, R. & Nizet, V. The ontogeny of a neutrophil: Mechanisms of granulopoiesis and homeostasis. *Microbiol. Mol. Biol. Rev.* **82**, 17. <https://doi.org/10.1128/MMBR.00057-17> (2018).

78. Fang, X., Duan, S. F., Gong, Y. Z., Wang, F. & Chen, X. L. Identification of key genes associated with changes in the host response to severe burn shock: A bioinformatics analysis with data from the gene expression omnibus (GEO) database. *J. Inflamm. Res.* **13**, 1029–1041. <https://doi.org/10.2147/JIR.S282722> (2020).
79. Zeqiraj, E. & Sicheiri, F. Getting a handle on glycogen synthase—Its interaction with glycogenin. *Mol. Asp. Med.* **46**, 63–69. <https://doi.org/10.1016/j.mam.2015.08.004> (2015).
80. Theilgaard-Mönch, K. *et al.* Haptoglobin is synthesized during granulocyte differentiation, stored in specific granules, and released by neutrophils in response to activation. *Blood* **108**, 353–361. <https://doi.org/10.1182/blood-2005-09-3890> (2006).
81. Yağcı, S., Serin, E., Acicbe, Ö., Zeren, M. & Odabaşı, M. S. The relationship between serum erythropoietin, hepcidin, and haptoglobin levels with disease severity and other biochemical values in patients with COVID-19. *Int. J. Lab. Hematol.* **43**(Suppl 1), 142–151. <https://doi.org/10.1111/ijlh.13479> (2021).
82. Kumar, N. P. *et al.* Systemic inflammation and microbial translocation are characteristic features of SARS-CoV-2-related multisystem inflammatory syndrome in children. *Open Forum Infect. Dis.* **8**, 279. <https://doi.org/10.1093/ofid/ofab279> (2021).
83. Yang, L. *et al.* Lectin microarray combined with mass spectrometry identifies haptoglobin-related protein (HPR) as a potential serologic biomarker for separating nonbacterial pneumonia from bacterial pneumonia in childhood. *Proteom. Clin. Appl.* **12**, e1800030. <https://doi.org/10.1002/prca.201800030> (2018).
84. Shin, E. C. *et al.* Expression patterns of alpha-synuclein in human hematopoietic cells and in *Drosophila* at different developmental stages. *Mol. Cells* **10**, 65–70. <https://doi.org/10.1007/s10059-000-0065-x> (2000).
85. Grozdanov, V. & Danzer, K. M. Intracellular alpha-synuclein and immune cell function. *Front. Cell Dev. Biol.* **8**, 562692. <https://doi.org/10.3389/fcell.2020.562692> (2020).
86. Barbut, D., Stolzenberg, E. & Zasloff, M. Gastrointestinal immunity and alpha-synuclein. *J. Parkinsons Dis.* **9**, S313–S322. <https://doi.org/10.3233/JPD-191702> (2019).
87. Rosen, B., Kurtishi, A., Vazquez-Jimenez, G. R. & Möller, S. G. The intersection of Parkinson's disease, viral infections, and COVID-19. *Mol. Neurobiol.* **58**, 4477–4486. <https://doi.org/10.1007/s12035-021-02408-8> (2021).
88. Miotto, M. *et al.* Molecular mechanisms behind anti SARS-CoV-2 action of lactoferrin. *Front. Mol. Biosci.* **8**, 607443. <https://doi.org/10.3389/fmolb.2021.607443> (2021).
89. Abers, M. S. *et al.* An immune-based biomarker signature is associated with mortality in COVID-19 patients. *JCI Insight* **6**, 44455. <https://doi.org/10.1172/jci.insight.144455> (2021).
90. Shrivastava, S. *et al.* Elevated levels of neutrophil activated proteins, alpha-defensins (DEFA1), calprotectin (S100A8/A9) and myeloperoxidase (MPO) are associated with disease severity in COVID-19 patients. *Front. Cell Infect. Microbiol.* **11**, 751232. <https://doi.org/10.3389/fcimb.2021.751232> (2021).
91. Zeng, Y., Ma, W., Ma, C., Ren, X. & Wang, Y. Inhibition of TXNDC5 attenuates lipopolysaccharide-induced septic shock by altering inflammatory responses. *Lab. Invest.* <https://doi.org/10.1038/s41374-021-00711-5> (2021).
92. Huang, D., Liu, A. Y. N., Leung, K. S. & Tang, N. L. S. Direct measurement of B lymphocyte gene expression biomarkers in peripheral blood transcriptomics enables early prediction of vaccine seroconversion. *Genes* **12**, 070971. <https://doi.org/10.3390/genes12070971> (2021).
93. Lopez-Castaneda, S. *et al.* Inflammatory and prothrombotic biomarkers associated with the severity of COVID-19 infection. *Clin. Appl. Thromb. Hemost.* **27**, 1076029621999099. <https://doi.org/10.1177/1076029621999099> (2021).
94. Sai, K. *et al.* Distal promoter regions are responsible for differential regulation of human orosomucoid-1 and -2 gene expression and acute phase responses. *Biol. Pharm. Bull.* **37**, 164–168. <https://doi.org/10.1248/bpb.b13-00551> (2014).
95. Liu, X. *et al.* Proteomics analysis of serum from COVID-19 patients. *ACS Omega* **6**, 7951–7958. <https://doi.org/10.1021/acsomega.1c00616> (2021).
96. Odle, R. I., Florey, O., Ktistakis, N. T. & Cook, S. J. CDK1, the other “master regulator” of autophagy. *Trends Cell Biol.* **31**, 95–107. <https://doi.org/10.1016/j.tcb.2020.11.001> (2021).
97. Auwal, M. R., Rahman, M. R., Gov, E., Shahjahan, M. & Moni, M. A. Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19. *Brief Bioinform.* **22**, 120. <https://doi.org/10.1093/bib/bbab120> (2021).
98. Malumbres, M. & Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* **30**, 630–641. <https://doi.org/10.1016/j.tibs.2005.09.005> (2005).
99. Yu, M. *et al.* Identification and validation of three core genes in p53 signaling pathway in hepatitis B virus-related hepatocellular carcinoma. *World J. Surg. Oncol.* **19**, 66. <https://doi.org/10.1186/s12957-021-02174-w> (2021).
100. Chen, J. C. *et al.* Identification of key pathways and genes in SARS-CoV-2 infecting human intestines by bioinformatics analysis. *Biochem. Genet.* <https://doi.org/10.1007/s10528-021-10144-w> (2021).
101. Gravina, G. *et al.* Survivin in autoimmune diseases. *Autoimmun. Rev.* **16**, 845–855. <https://doi.org/10.1016/j.autrev.2017.05.016> (2017).
102. Pulkkinen, H. H. *et al.* BMP6/TAZ-Hippo signaling modulates angiogenesis and endothelial cell response to VEGF. *Angiogenesis* **24**, 129–144. <https://doi.org/10.1007/s10456-020-09748-4> (2021).
103. Jiang, P. *et al.* Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat. Methods* **18**, 1181–1191. <https://doi.org/10.1038/s41592-021-01274-5> (2021).
104. Liu, J. *et al.* The role of 14–3–3 proteins in cell signalling pathways and virus infection. *J. Cell Mol. Med.* **25**, 4173–4182. <https://doi.org/10.1111/jcmm.16490> (2021).
105. Vieira, S. E. *et al.* Distinct transcriptional modules in the peripheral blood mononuclear cells response to human respiratory syncytial virus or to human rhinovirus in hospitalized infants with bronchiolitis. *PLoS ONE* **14**, e0213501. <https://doi.org/10.1371/journal.pone.0213501> (2019).
106. Girkin, J. L. N., Maltby, S. & Bartlett, N. W. Toll-like receptor-agonist-based therapies for respiratory viral diseases: Thinking outside the cell. *Eur. Respir. Rev.* <https://doi.org/10.1183/16000617.0274-2021> (2022).
107. Ramasamy, R. Innate and adaptive immune responses in the upper respiratory tract and the infectivity of SARS-CoV-2. *Viruses* **14**, 50933. <https://doi.org/10.3390/v14050933> (2022).
108. Wylezich, C. *et al.* Whole-genome analysis of SARS-CoV-2 samples indicate no tissue specific genetic adaptation of the virus in COVID-19 patients' upper and lower respiratory tract. *Diagn. Microbiol. Infect. Dis.* **101**, 115520. <https://doi.org/10.1016/j.diagmicrobio.2021.115520> (2021).
109. Laforge, M. *et al.* Tissue damage from neutrophil-induced oxidative stress in COVID-19. *Nat. Rev. Immunol.* **20**, 515–516. <https://doi.org/10.1038/s41577-020-0407-1> (2020).
110. Voinsky, I., Baristaite, G. & Gurwitz, D. Effects of age and sex on recovery from COVID-19: Analysis of 5769 Israeli patients. *J. Infect.* **81**, e102–e103. <https://doi.org/10.1016/j.jinf.2020.05.026> (2020).
111. Dhochak, N., Singhal, T., Kabra, S. K. & Lodha, R. Pathophysiology of COVID-19: Why children fare better than adults? *Indian J. Pediatr.* **87**, 537–546. <https://doi.org/10.1007/s12098-020-03322-y> (2020).
112. Garcia, G. *et al.* Hippo signaling pathway activation during SARS-CoV-2 infection contributes to host antiviral response. *BioRxiv.* <https://doi.org/10.1101/2022.04.07.487520> (2022).
113. Gao, Y. D. *et al.* Risk factors for severe and critically ill COVID-19 patients: A review. *Allergy* **76**, 428–455. <https://doi.org/10.1111/all.14657> (2021).

## Acknowledgements

The authors would like to thank Dr. Daniel Santa Cruz Damineli (post-doctoral fellow, Grant FAPESP No. 2019/23343-7), for valuable help on data analysis; Roberta Cristiane Pascharelli Alves and Franscisca Laudeci Felix Bezerra for collecting biological samples from oligosymptomatic individuals. This research was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Grants Nos. 2015/22308-2, 2019/23343-7, and 2020/06160-3 to C.A.M.-F and R.M.C.J.; and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grant No. 306893/2018-5 to C.A.M.F.

## Author contributions

Conceived and designed the experiments: S.E.V., C.A.M.F., S.Y.B. Performed the experiments: S.Y.B., F.B.B., S.E.V., D.B.L.O., V.N.C., E.L.D., P.P., A.C.P.C., C.S.F., G.P.L., F.A.R., C.A.M.F. Analyzed the data: S.Y.B., F.B.B., S.E.V., D.B.L.O., E.L.D., P.P., C.A.M.F. Funding acquisition: C.A.M.F., E.L.D., R.M.C.J. Resources: E.L.D., P.P., L.A., G.P.L., R.M.C.J., C.A.M.F. Wrote the paper: S.Y.B., F.B.B., S.E.V., D.B.L.O., E.L.D., C.A.M.F. All authors have read and agreed to the published version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-28227-6>.

**Correspondence** and requests for materials should be addressed to C.A.M.-F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2023