Multi-layered transcriptomic analyses reveal an immunological overlap between COVID-19 and hemophagocytic lymphohistiocytosis associated with disease severity

Lena F. Schimke\textsuperscript{a,5}, Alexandre H.C. Marques\textsuperscript{b}, Gabriela Crispim Baiocchi\textsuperscript{a}, Caroline Aliane de Souza Prado\textsuperscript{b}, Dennysn Leandro M. Fonseca\textsuperscript{b}, Paula Pacciarelli Freire\textsuperscript{d}, Desirée Rodrigues Plaça\textsuperscript{b}, Igor Salerno Filgueiras\textsuperscript{a}, Ranieri Coelho Salgado\textsuperscript{a}, Gabriel Jansen-Marques\textsuperscript{c}, Antonio Edson Rocha Oliveira\textsuperscript{b}, Jean Pierre Schatzmann Peron\textsuperscript{a}, José Alexandre Marzagão Barbuto\textsuperscript{a,d}, Niels Olsen Saraiva Camarã\textsuperscript{a}, Vera Lúcia Garcia Calich\textsuperscript{a}, Hans D. Ochs\textsuperscript{e}, Antonio Condino-Neto\textsuperscript{b}, Katherine A. Overmyer\textsuperscript{f,g}, Joshua J. Coon\textsuperscript{h,i}, Joseph Balnis\textsuperscript{j,k}, Ariel Jaitovich\textsuperscript{j,k}, Jonas Schulte-Schrepping\textsuperscript{l}, Thomas Ulas\textsuperscript{m}, Joachim L. Schultz\textsuperscript{l,m}, Helder I. Nakaya\textsuperscript{a}, Igor Jurisica\textsuperscript{n,o,p}, Otavio Cabral-Marques\textsuperscript{a,b,q}

\textsuperscript{a}Departamento de Imunologia, Instituto de Ciencias Biomedicas, Universidade de São Paulo, São Paulo, SP, Brazil. ID Ringgold USP + ICB: 54544, ISNI: 0000000406355304
\textsuperscript{b}Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, SP, Brazil.
\textsuperscript{c}Information Systems, School of Arts, Sciences and Humanities, University of Sao Paulo, São Paulo, SP, Brazil.
\textsuperscript{d}Laboratory of Medical Investigation in Pathogenesis and targeted therapy in Onco-Immuno-Hematology (LIM-31), Department of Hematology, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil.
\textsuperscript{e}National Center for Quantitative Biology of Complex Systems, Madison, WI 53562, USA
\textsuperscript{f}Morgridge Institute for Research, Madison, WI 53562, USA
\textsuperscript{g}Department of Pulmonary and Critical Care Medicine, Albany Medical Center, Albany, NY 12208, USA
\textsuperscript{h}Department of Molecular and Cellular Physiology, Albany Medical College, Albany, NY 12208, USA
\textsuperscript{i}Life and Medical Sciences (LIMES) Institute, University of Bonn, Germany.
\textsuperscript{j}German Center for Neurodegenerative Diseases (DZNE), PRECISE Platform for Genomics and Epigenomics at DZNE, and University of Bonn, Bonn, Germany.
\textsuperscript{k}Osteoarthritis Research Program, Division of Orthopedic Surgery, Schroeder Arthritis Institute and Data Science Discovery Centre for Chronic Diseases, Krembil Research Institute, University Health Network, Toronto, Canada.
\textsuperscript{l}Departments of Medical Biophysics and Computer Science, University of Toronto, Toronto, Canada.
\textsuperscript{m}Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovakia.
\textsuperscript{n}Network of Immunity in Infection, Malignancy, and Autoimmunity (NIIMA), Universal Scientific Education and Research Network (USERN), Sao Paulo, Brazil.
\textsuperscript{5}Lead contact

Corresponding authors:
Lena F. Schimke, MD
Department of Immunology
Institute of Biomedical Sciences - University of São Paulo
Lineu Prestes Avenue, 1730,
São Paulo, SP, CEP 05508-900, Brazil
e-mail: lenaschimke@hotmail.com
phone.: +55 (11) 3091-7387

Otavio Cabral-Marques, MSc, PhD
Department of Immunology
Institute of Biomedical Sciences - University of São Paulo
Lineu Prestes Avenue, 1730,
São Paulo, SP, CEP 05508-900, Brazil
e-mail: otavio.cmarques@usp.br
phone: +55 11 974642022
ABSTRACT

Clinical and hyperinflammatory overlap between COVID-19 and hemophagocytic lymphohistiocytosis (HLH) has been reported. However, the underlying mechanisms are unclear. Here we show that COVID-19 and HLH have an overlap of signaling pathways and gene signatures commonly dysregulated, which were defined by investigating the transcriptomes of 1253 subjects (controls, COVID-19, and HLH patients) using microarray, bulk RNA-sequencing (RNAseq), and single-cell RNAseq (scRNAseq). COVID-19 and HLH share pathways involved in cytokine and chemokine signaling as well as neutrophil-mediated immune responses that associate with COVID-19 severity. These genes are dysregulated at protein level across several COVID-19 studies and form an interconnected network with differentially expressed plasma proteins which converge to neutrophil hyperactivation in COVID-19 patients admitted to the intensive care unit. scRNAseq analysis indicated that these genes are specifically upregulated across different leukocyte populations, including lymphocyte subsets and immature neutrophils. Artificial intelligence modeling confirmed the strong association of these genes with COVID-19 severity. Thus, our work indicates putative therapeutic pathways for intervention.

Keywords:

COVID-19, Hemophagocytic lymphohistiocytosis, common signaling pathways, gene signatures
INTRODUCTION

More than one year of Coronavirus disease 2019 (COVID-19) pandemic caused by the severe acute respiratory syndrome Coronavirus (SARS-CoV)-2, more than 197 million cases and 4.2 million deaths have been reported worldwide (July 30th 2021, WHO COVID-19 Dashboard). The clinical presentation ranges from asymptomatic to severe disease manifesting as pneumonia, acute respiratory distress syndrome (ARDS), and a life-threatening hyperinflammatory syndrome associated with excessive cytokine release (hypercytokinaemia). Risk factors for severe manifestation and higher mortality include old age as well as hypertension, obesity, and diabetes. Currently, COVID-19 continues to spread, new variants of SARS-CoV-2 have been reported and the number of infections resulting in death of young individuals with no comorbidities has increased the mortality rates among the young population. In addition, some novel SARS-CoV-2 variants of concern appear to escape neutralization by vaccine-induced humoral immunity. Thus, the need for a better understanding of the immunopathologic mechanisms associated with severe SARS-CoV-2 infection.

Patients with severe COVID-19 have systemically dysregulated innate and adaptive immune responses, which are reflected in elevated plasma levels of numerous cytokines and chemokines including granulocyte colony-stimulating factor (GM-CSF), tumor necrosis factor (TNF), interleukin (IL)-6, IL-6R, IL18, CC chemokine ligand 2 (CCL2) and CXC chemokine ligand 10 (CXCL10), and hyperactivation of lymphoid and myeloid cells. Notably, the hyperinflammation in COVID-19 shares similarities with cytokine storm syndromes such as those triggered by sepsis, autoinflammatory disorders, metabolic conditions and malignancies, often resembling a hematopathologic condition called hemophagocytic lymphohistiocytosis (HLH). HLH is a life-threatening progressive systemic hyperinflammatory disorder characterized by multi-organ involvement, fever flares, hepatosplenomegaly, and cytopenia due to hemophagocytic activity in the bone marrow or within peripheral lymphoid organs such as pulmonary lymph nodes and spleen. HLH is marked by aberrant activation of B and T lymphocytes and monocytes/macrophages, coagulopathy, hypotension, and ARDS. Recently, neutrophil hyperactivation has been shown to also play a critical role in HLH development. This is in agreement with the observation that the HLH-like phenotype observed in severe COVID-19 patients is due to an innate neutrophilic hyperinflammatory response associated with...
virus-induced hypercytokinaemia which is dominant in patients with an unfavorable clinical course\textsuperscript{17}. Thus, HLH has been proposed as an underlying etiologic factor of severe COVID-19\textsuperscript{1,3,20}. HLH usually develops during the acute phase of COVID-19\textsuperscript{1,20–27}. However, a case of HLH that occurred two weeks after recovery from COVID-19 has recently been reported as the cause of death during post-acute COVID-19 syndrome\textsuperscript{28}.

The familial form of HLH (fHLH) is caused by inborn errors of immunity (IEI) in different genes encoding proteins involved in granule-dependent cytotoxic activity of leukocytes such as AP3B1, LYST, PRF1, RAB27A, STXBP2, STX11, UNC13D\textsuperscript{29–31}. In contrast, the secondary form (sHLH) usually manifests in adults following a viral infection (e.g., adenovirus, EBV, enterovirus, hepatitis viruses, parvovirus B19, and HIV)\textsuperscript{32,33}, or in association with autoimmune/rheumatologic, malignant, or metabolic conditions that lead to defects in T/NK cell functions and excessive inflammation\textsuperscript{16,31}. fHLH and sHLH affect both children and adults, however, the clinical and genetic distinction of HLH forms is not clear since immunocompetent children can develop sHLH\textsuperscript{34,35}, while adult patients with sHLH may also have germline mutations in HLH genes\textsuperscript{36}. Of note, germline variants in UNC13D and AP3B1 have also been identified in some COVID-19 patients with HLH phenotype\textsuperscript{37}, thus, indicating that both HLH forms may be associated with COVID-19.

Here, we characterized the signaling pathways and gene signatures commonly dysregulated in both COVID-19 and HLH patients by investigating the transcriptomes of 1253 subjects (controls, COVID-19, and HLH patients) assessed by microarray, bulk RNA-sequencing (RNAseq), and single-cell RNAseq (scRNAseq) (Table 1). We found shared gene signatures and cellular signaling pathways involved in cytokine and chemokine signaling as well as neutrophil-mediated immune responses that associate with COVID-19 severity.
RESULTS

The transcriptional overlap between COVID-19 and HLH

Considering the fact that COVID-19 and HLH\textsuperscript{14,21} share clinical and biological features (Figure 1A) we suspected that these diseases have a transcriptional overlap underlying their phenotypic convergence. First, we obtained differentially expressed genes (DEGs) for each dataset from different high-throughput transcriptome technologies present in peripheral blood lymphocytes (PBLs), peripheral blood mononuclear cells (PBMCs), and nasopharyngeal swabs from COVID-19 patients, HLH patients and controls (Figure 1B and Supp. Table S1). Then, we defined the transcriptome overlap between DEGs from COVID-19 cohorts and HLH by cross-technology comparisons, performed enrichment analysis and association studies between specific DEGs and severity status of disease (Figure 1C). To identify the common DEGs we divided the datasets into three subgroups based on type of samples and RNAseq platforms: Overlap 1 (HLH and COVID-19 blood transcriptomes), Overlap 2 (HLH and COVID-19 nasopharyngeal swab transcriptomes), and Overlap 3 (HLH and COVID-19 scRNAseq transcriptomes) (Supp. Figure 1A and 1B, and Supp. Table S2 and S3). We found a total of 239 unique common DEGs between HLH and all COVID-19 datasets, most of them (237 DEGs) up-regulated (Figure 1D). Hereafter, we focused on the implications of the up-regulated genes, since the 2 common down-regulated genes (granulysin or \textit{GNLY}; myomesin 2 or \textit{MYOM2}) alone did not enrich any significant pathway. However, this might also indicate a defect in cytotoxicity activity, typical of HLH\textsuperscript{31}, that will require future investigation. The 237 common up-regulated DEGs encode proteins mainly involved in immune system, metabolic and signaling processes, forming a highly connected biological network based on physical protein-protein interactions (PPI, Figure 1E). Among them are important genes encoding molecules involved in activation of inflammatory immune responses (e.g., PGLYRP1, OLR1, FFAR2), cytokine and chemokine mediated immune pathways (e.g., IL1R2, CXCR2, CXCR8, CCL4, CCL2), and neutrophil activation (e.g., CD177, MPO, ELANE). Of note, the transcriptional overlap between HLH and COVID-19 contains several molecules interacting with 7 genes causing fHLH due to IEL which itself were not among our DEGs (Figure 1E).
Cytokine/chemotaxis and neutrophil signatures predominate in COVID-19 and HLH

We next dissected the biological functions enriched by the 237 common up-regulated DEGs between COVID-19 and HLH patients by performing enrichment analysis of biological processes (BPs) and cellular components (CCs) by these 237 DEGs. The top 20 most enriched BPs are demonstrated in Figure 2A, which encompass cytokine/chemotaxis and neutrophil-mediated innate immune responses, ranging from response to IL-1 to neutrophil activation, degranulation, and migration (for all BPs see Supp. Table S4). The CCs enriched (Figure 2B) include several compartments such as secretory granule lumen and membrane, azurophil tertiary and specific granules, as well as collagen-containing extracellular matrix, phagocytic vesicle, and primary lysosome (Supp. Table S5).

Cytokine/chemotaxis and neutrophil signatures predominate in the COVID-19 and HLH multi-layered transcriptional overlap. A total of 25, 34, and 58 DEGs are assigned to cytokine, chemotaxis, and neutrophil signatures, respectively (Figure 2C-E: complete categorization can be seen in Supp. Table S6 and S7). Several genes play pleiotropic roles in these gene ontology (GO) categories such as CEACAM8, IL-1β, IL-6, EDN1, NFKB1 and PDE4B (Supp. Table S8). For clarity in data visualization, we assigned these genes to a unique category (based on their predominant immunological function according to literature and GeneCards; the human gene database). Among these genes, there are chemokines and chemokine receptors that attract both lymphocytes and myelocytes to inflammation sites (CCL20, CCL2, CXCR1, CXCR3, CXCL8)\(^{39,40}\), pro-inflammatory cytokines (IL-1B, IL-1R1, NFKB1, IFNG, IL-6, TNF)\(^{41,42}\) that promote the activation of immune cells, and several proteins/granules with antimicrobial activity (MPO\(^{43}\), AZU1\(^{44}\), ELANE\(^{45,46}\), DEFA4\(^{47}\)). Moreover, there are metalloproteinases (MMP8 and MMP9) involved in degradation of extracellular matrix (ECM) for neutrophil migration\(^{8,49}\) into the airways and in the regulation of cytokine activity. Of note, hierarchical clustering analysis of these genes indicated a cross-study grouping of closer functional-related molecules. For instance, IFNG, IL6, and TNF; IL1A and IL1B; as well as signaling molecules involved in the nuclear factor-κB (NF-κB) such as NFKB1, SPHK1, and RIPK2 clustered together in the cytokine group. Likewise, GYPB and GYPB; CEACAM6 and CEACAM8; as well as CCL and CXCL chemokines...
in the chemotaxis group and antimicrobial-related peptides such as *AZU1, MPO, CAMP, DEFA4, LCN2, ELANE, OLFM4,* and *CD177* in neutrophil-mediated immunity. However, we cannot exclude that this clustering pattern just represent an aleatory tendency due to upstream GO categorization.

Of note, the classical literature has emphasized that the 7 genes (*AP31B, LYST, PRF1, RAB27A, STX11, STXBP2, UNC13D*) known to cause fHLH (classically defined as familial HLH syndromes and hypopigmentation syndromes)\(^\text{14,29}\) suggest a dominant role for T and NK cells in the development of HLH\(^\text{50–52}\). However, these genes also enrich several CCs (secretory vesicles, azurophilic granules or specific granules) and BPs (neutrophil degranulation) involved in the neutrophil immune responses (Supp. Figure 2). This result is in agreement with the role of these genes in a variety of neutrophil functions such as degranulation and formation of neutrophil extracellular traps (NET)\(^\text{53–57}\). Altogether, these data indicate gene signatures and signaling pathways commonly dysregulated in COVID-19 and HLH patients that involve a network of regulatory (production of chemokines/cytokines by T cells and macrophages) and effector (neutrophil hyperactivation, NK and T cell cytotoxicity) immune functions.

The relationship between cytokine/chemotaxis and neutrophil-mediated immunity gene signatures

We next analyzed the relationship pattern and degree between the transcriptional signatures related to cytokine signaling and chemotaxis with those involved in neutrophil-mediated immune responses. We chose the COVID-19_PBL dataset from Overmyer et al.\(^\text{58}\), which contains transcripts from 100 individuals with COVID-19 and 26 individuals with respiratory symptoms but negative for COVID-19 serving as control group (further explored in the next session). We performed canonical-correlation analysis (CCA), which is a multivariate statistical method to determine the linear relationship between two groups of variables\(^\text{59}\). In accordance with the cross-study hierarchical clustering, CCA revealed a strong association between several cytokine/chemotaxis related genes (e.g., *CXCL8, CEACAMs [1/6/8], IL1RAP, IL1R1, IL1B, NFKB1*) with those involved in neutrophil-mediated immune responses (e.g., *CTSG, ELANE, MMP8, TCN1*) in both patients with COVID-19 and controls (Figure 3A and 3B). Bivariate correlation
analysis showed a similar phenomenon (Supp. Figure 3). However, these correlation patterns partially changed when comparing COVID-19 with the control group. For instance, while reducing the correlation between molecules including IL-10, CXCL8, NFKB1, ARG1, and SOD2, new strong associations appeared between ELANE, DEFA4, AZU1, CTSG, and LCN2, with an overall tendency to higher relationships amid neutrophil-mediated immunity related genes in COVID-19 patients. Figure 3C illustrates this observation by scatter plots for some of these variables.

Transcripts stratifying severe COVID-19 from other respiratory diseases are also highly dysregulated in HLH

Next, we sought to determine which genes of chemotaxis/cytokine signaling and neutrophil-mediated immune responses discriminate COVID-19 patients according to disease severity. We further investigated the COVID-19_PBL dataset (GSE157103) comparing COVID-19 patients admitted to the intensive care unit (COVID-19_ICU) with those admitted to non-ICU units (COVID-19_nonICU). Severity of critical illness at ICU admission was defined based on APACHE II and SOFA scores according to Overmyer et al. (Figure 4A). Among all genes, 25 (15 up-regulated and 10 down-regulated genes) were differentially expressed between COVID-19_ICU and COVID19_nonICU patients (Supp. Table S9). These 25 genes were also present and up-regulated in peripheral blood leukocytes of COVID-19 patients from two recently published works, which were unpublished when we started our investigation (GSE163151 and GSE152641, Supp Table S10). Of note, most of these 25 genes have also been identified at protein level as dysregulated in COVID-19 patients across different studies (published during the development of our study; Supp. Table S11). In addition, these 25 genes might belong to a systemic immune network of molecules induced by SARS-CoV-2 since they are also highly interconnected with 158 proteins (Supp. Table 12) significantly dysregulated in the plasma of COVID-19_ICU when compared to COVID-19_nonICU patients. Thus, they show several interactions and functional overlap (Figure 4B) with plasma proteins involved in neutrophil degranulation and neutrophil-mediated immunity (Supp. Figure 4).
To investigate the stratification power of these 25 DEGs, we performed principal component analysis (PCA) using a spectral decomposition approach\textsuperscript{63,64}, which examines the covariances/correlations between variables. This approach revealed that these DEGs clearly divide COVID-19\_ICU, COVID-19\_nonICU, Control\_ICU and Control\_nonICU (due to other respiratory illness but negative for SARS-CoV-2) groups (Figure 4C and Supp. Figure 5A and 5C). Likewise, these 25 genes stratified HLH patients from healthy controls (Figure 4D and Supp. Figure 5B and 5D). As seen in these PCA graphics, 8 (AZU1, CEACAM8, CTSG, DEFA4, ELANE, LCN2, OLFM4, and MMP8) of these 25 DEGs strongly associate with COVID-19\_ICU. Bivariate correlation analysis based on these 25 genes showed that while controls and COVID-19\_nonICU patients have a similar general cluster distribution, COVID-19\_ICU patients tend to differ of them, maintaining only the 8 genes with high positive correlations (Figure 4E and Figure 4F). Combined, these results suggest that these genes are associated with COVID-19 severity and highlight the importance of neutrophil-mediated immunity related molecules in severe COVID-19 on both transcriptomic and proteomic level.

**Multi-layered transcriptomic analysis associates COVID-19 and HLH common genes with disease severity**

Since scRNAseq allows comparison of the transcriptomes of individual cells, we next sought to investigate the distribution patterns of these 25 genes associated with COVID19 severity. We analyzed the scRNAseq dataset (EGAS00001004571) reported by Schulte-Schrepping et al.\textsuperscript{65} (schematic overview of study group Figure 5A) and found that 21 of the 25 genes associated with COVID-19 severity and HLH development are DEGs among the top 2,000 variable genes in the COVID-19 cohort compared to controls (Figure 5B and Supp. Fig 6A and B). These 21 genes exhibited cell-type-specific expression patterns. For instance, CCL4 (a chemoattractant and stimulator of T-cell immune responses\textsuperscript{66,67}) was mainly produced by CD8+ T and NK cells, CD83 (B, T and dendritic cell activation marker\textsuperscript{68,69}) by B cells and monocytes. CXCL8 was mostly present in monocytes and low-density neutrophils/granulocytes (LDGs; also frequently reported as immature neutrophils\textsuperscript{70–72}), which are neutrophils remaining in the PBMC fraction after density gradient separation. Among these 21 genes, 11 genes (among them also the 8 genes
described above) were differentially expressed when comparing patients with mild and severe COVID-19 (Figure 5C). Thus, indicating a network of cell-type-specific expression patterns that may contribute to the clinical similarities between COVID-19 and HLH.

Of note, these 11 genes encode proteins that are crucial for several pathways involved in neutrophil-mediated immunity, and are associated with diseases that increase the risk of severe COVID-19 such as chronic obstructive pulmonary disease (COPD) and ulcerative colitis (Supp. Figure 6E and Supp. Table S14). Similarly, these 11 genes are also significantly different between COVID-19_ICU and COVID-19_nonICU (Figure 5E) in the bulk RNAseq dataset (GSE157103, Overmyer et al. 2020). Indicating that these genes are consistently associated with COVID-19 severity across different patient cohorts. Moreover, these 11 genes were differentially expressed in HLH patients compared to healthy controls (Figure 5D), reinforcing the transcriptomic overlap of severe COVID-19 and HLH.

We used random forest method to rank the importance of these 11 genes based on their ability to discriminate between COVID-19_ICU and COVID-19_nonICU, hence evaluating the association of these genes with the COVID-19 severity. This approach showed error rate (out of bag or OOB) of 27.03% and an area under the ROC curve of 82.4% for both groups (Figure 6A and 6B). Follow-up analysis indicated that ARG1 was the most significant predictor for ICU admission followed by CD177, MCEMP1, LCN2, AZU1, OLFM4, MMP8, ELANE, CTSG, DEFA4, CEACAM8 based on the number of the nodes, gini-decrease, and average depth criteria for measuring gene importance (Figure 6C and 6D). ARG1 exhibited the most relevant interactions with the other genes according to the mean minimal depth criterion, mostly interacting with CD177, AZU1, MCEMP1, and LCN2 (Figure 6D). Altogether, these multi-layered transcriptomic results associate COVID-19 and HLH common genes with disease severity.

DISCUSSION

The results of our study provide a COVID-19/HLH immune landscape using a multi-layered transcriptomic approach of microarray, bulk and scRNAseq. Our meta-analysis integrates and unravels the consistency of several important individual studies and datasets that also validated the transcriptome data at the protein level in COVID-19 patients. To the best of our
knowledge, it represents the first attempt to systems characterize the common signaling pathways and molecular networks shared by COVID-19 and HLH. In agreement with the recent observation that neutrophil hyperactivation plays a key role in the severity of COVID-19 and HLH\textsuperscript{18,19}, our approach indicates that COVID-19 and HLH have a common transcriptional profile formed predominantly by a group of regulatory molecules related to cytokine/chemotaxis and by a group of effector molecules that are linked to neutrophil hyperactivation and disease severity. These data highlight the dual role of neutrophils in providing essential antimicrobial functions, but also initiating tissue injury caused by immune dysregulation\textsuperscript{85,86}. The genes associated with COVID-19 severity are up-regulated across different leukocyte subpopulations such as lymphoid (NK, T and B cells) and myeloid (monocytes, dendritic cells and LDGs) cells. They form a systemic and interconnected network of cell-type-specific expression pattern and signaling networks that may contribute to the clinical similarities between COVID-19 and HLH. Thus, our analysis identified new candidate biomarkers and novel putative molecular pathways for therapeutic intervention for COVID-19 and HLH.

Our work expands the efforts of others\textsuperscript{87–91} and our group\textsuperscript{92} to identify networks and pathways involved in the pathogenesis of severe COVID-19. Our multi-layered transcriptomics approach is in agreement with the computational model developed by Ding et al.\textsuperscript{18}, which is based on a network-informed analysis of the interaction of SARS-CoV-2 and HLH related genes. Their model postulates that neutrophil degranulation/neutrophil extracellular traps (NETs) cause endothelial damage, and consequently, thrombotic complications of COVID-19. Ding’s and our interpretation is supported by several experimental evidence\textsuperscript{81–83,93} for neutrophil hyperactivation and its association with severity of COVID-19, as recently reviewed by Ackermann et al.\textsuperscript{84}. As we were able to demonstrate by the multi-omics association between leukocyte and plasma molecules, flow cytometry and proteomic data indicate a systemic and integrated network of molecules associated with neutrophil growth, activation, and mobilization leading to neutrophil dysregulation in severe COVID-19\textsuperscript{82,83}. This supports the concept that the pathophysiology of HLH does not only involve T cell, NK cell and macrophage dysregulation but also the hyperactivation of neutrophils. In accordance with our findings, it has recently been demonstrated that neutrophils accumulate in inflamed tissues of HLH and COVID-19 as a consequence of T-cell driven pro-inflammatory cytokine and chemokine release, which
does not return to a homeostatic level due to an ineffective T cytotoxic response\(^{18,19}\). Thus, maintaining the immunogenic stimulus for more cytokine/chemokine secretion, which promotes the sustained neutrophil recruitment and consequent tissue damage.

The use of high-throughput techniques to identify biomarkers, molecular pathways and pathophysiological information derived from genetic and transcriptomic data has contributed to the understanding of the immunopathology of diseases\(^94\). In this regard, patients with severe COVID-19 have been associated with mutations in genes involved in the regulation of type I and III IFN immunity pointing to the role of structural genomics in determining the course of COVID-19\(^95\). Our functional genomics approach characterized by a multi-layered transcriptomic analysis started with dozens of thousand genes expressed in nasopharyngeal swabs and peripheral blood obtained from 1253 subjects. We identified 25481 DEGs, of which 237 were commonly up-regulated in COVID-19 and HLH. Among them, 25 genes were differentially expressed in COVID-19\_ICU in comparison to COVID-19\_nonICU, demonstrating the disruption in gene network signatures and reinforcing the concept that alterations in molecule relationships are immunopathological mechanisms also involved in the breakdown of body homeostasis\(^96\). In this context, 11 genes (\textit{ARG1}, \textit{AZU1}, \textit{CD177}, \textit{CEACAM8}, \textit{CTSG}, \textit{DEFA4}, \textit{ELANE}, \textit{LCN2}, \textit{MCEMP1}, \textit{MMP8}, and \textit{OLFM4}) commonly dysregulated in COVID-19 and HLH specifically stratified COVID-19\_ICU from COVID-19\_nonICU patients. They encode proteins involved in neutrophil degranulation and contribute to the development of comorbidities that increase the risk of progressing to severe COVID-19\(^73,74\). Random forest model ranking indicated that these genes accurately distinguish COVID-19\_nonICU from COVID-19\_ICU patients. For instance, this machine learning approach ranked \textit{ARG1} and its interaction with other molecules (\textit{CD177}, \textit{AZU1}, \textit{MCEMP1} and \textit{LCN2}) as an important predictor for ICU admission, supporting the role of these molecules as biomarkers for severe COVID-19\(^80,97\).

It is important to mention that while our work has the strengths of a robust cross-study and cross-tissue network analysis, it has limited value to mechanistically explore the role of specific molecules in the outcome of COVID-19. However, several of these dysregulated molecules shared by COVID-19 and HLH have been successfully investigated for the treatment of SARS-CoV-2 infection, supporting our findings. For instance, inhibition of the CCR5-CCL4 axis by Leronlima (anti-CCR5 monoclonal antibody)\(^98\), or blockade of cytokine signaling by
Tocilizumab (anti-IL-6R)\textsuperscript{99}, Adalimumab (anti-TNF)\textsuperscript{100}, or Anakinra (anti-IL1R)\textsuperscript{101} have been shown to ameliorate, in some cases, severe COVID-19. Furthermore, Ruxolitinib, a JAK1/JAK2 inhibitor acting downstream of JAK-dependent chemokines/cytokines such as IFN-γ, IL-1β, IL-6, TNF, G-CSF, CXCL9, and CXCL10\textsuperscript{19,102} has shown promising results in treating COVID-19\textsuperscript{103}. Of note, several approaches targeting neutrophils to treat SARS-CoV-2 complications have entered clinical trials, including the disruption of signaling via CXCR2, IL-8, IL-17A, or the use of phosphodiesterase (PDE) inhibitors\textsuperscript{104}. Moreover, in agreement with our data, inhibition of neutrophil-derived anti-microbial proteins are being actively investigated in clinical trials by exploring the mechanistic and clinical effects of alvelestat, an oral neutrophil elastase inhibitor (COVID-19 Study of Safety and Tolerability of Alvelestat, ClinicalTrials.gov). Meanwhile other proteases (AZU1 and CTSG) and the inhibition of NETs have been suggested to alleviate SARS-CoV-2 symptoms\textsuperscript{80,84,105}.

In conclusion, our comprehensive multi-layered transcriptomic and cross-tissue analysis of nasopharyngeal swabs and peripheral blood leukocytes indicates systemic communalities among severe COVID-19 and in HLH. Our functional genomic work shows that the overlapping molecular pathways between COVID-19 and HLH are characterized by an interconnected cytokine/chemokine profile that hyper stimulates and systemically attracts adaptive and innate immune cells, culminating in the hyperactivation of neutrophils and leading to life-threatening clinical complications. Thus, our work suggests novel putative molecular pathways that can be exploited for therapeutic intervention for both COVID-19 and HLH.
ACKNOWLEDGMENTS

We acknowledge the Latin American Society of Immunodeficiencies (LASID) for providing the research funding of LFS (LASID Fellowship award 2020), and the São Paulo Research Foundation (FAPESP grants 2018/18886-9, 2020/01688-0, and 2020/07069-0 to OCM) for financial support. Computational analysis was supported by FAPESP and partially by the grants from Ontario Research Fund (#34876), Natural Sciences Research Council (NSERC #203475), Canada Foundation for Innovation (CFI #29272, #225404, #33536), and IBM granted to IJ, as well as the National Institutes of Health (NHLBI) through award HL130704 granted to AJ. This study was financed in part by the coordination for the improvement of higher education personnel – Brazil (CAPES) – finance code 001. We thank Prof. Luis Carlos de Souza Ferreira for discussion and suggestions to develop this project.

AUTHOR CONTRIBUTIONS

LFS, and OCM wrote the manuscript, conceived the study design, bioinformatics analyses and performed scientific insights; LFS, AHCM, GCB, CASP, DLMF, PPF, DRP, ISF, RCS, GJM, AERO, and IJ, co-wrote the manuscript and performed bioinformatics analyses; KAO, JJC, JB, AJ, JSS, TU and JLS generated transcriptome and proteome data; JPSP, JAMB, NOSC, HDO, VLGC, ACN, KAO, JJC, JB, AJ, JSS, TU, JLS, HIN, IJ and OCM revised and edited the final manuscript and provided scientific input; OCM supervised the project.

COMPETING INTEREST STATEMENT

The authors declare no competing financial and/or non-financial interests concerning the work described.
FIGURE LEGENDS:

Graphic Abstract: Schematic view of datasets and results obtained indicating new candidate biomarkers and therapeutic targets for COVID-19 and HLH (Figure created using BioRender.com).

Figure 1: Transcriptional overlap between COVID-19 and HLH. (A) Phenotypic convergence between severe COVID-19 and HLH. (B) Number of differentially expressed genes (DEGs, up- and down-regulated) by dataset. (C) Study design for multi-transcriptomics analysis. (D) Circos plot showing 237 common up-regulated DEGs (or transcripts) between HLH and the different COVID-19 datasets (red lines), divided in 3 overlapping subgroups (Detailed in Supp. Table 2 and 3). The thickness of each line represents the number of genes shared between the different datasets. (E) Protein-protein interaction network among the 237 transcripts and the 7 genes causing HLH due to inborn errors of immunity (IEI). Node colours denote Gene Ontology biological processes. The label (gene name) colours represent transcripts from Overlap 1 (green), Overlap 2 (red), and Overlap 3 (blue). Center circle and side circles represent common molecules across all 3 or 2 overlapping datasets, respectively. The upper left subnetwork represents the interactions between the 7 genes associated with fHLH and those from overlaps are bold. The circle on upper left (gene names not shown) contains 1329 proteins connected by 217 interactions with the 7 HLH/IEI-associated genes. The full network comprises 1538 proteins and 2522 direct physical interactions obtained from IID database ver. 2021-05

Figure 2: Cytokine/chemotaxis and neutrophil-associated transcriptional signatures predominate in the COVID-19 and HLH overlap. (A) Dot plot showing the most significant biological processes enriched by the 237 common up-regulated transcripts of COVID-19 and HLH datasets. The dot size is proportional to the number of genes enriching the ontology term and color proportional to adjusted p value (green > significant than blue). (B) Network highlighting genes and cellular component associations. Only enriched terms with adjusted p value <0.05 are shown. The degree of associations is displayed by edge color and thickness (e.g., lighter color and thinner edges signify fewer connections). Node color represents different GO CCs. Both enriched CCs and BPs were analyzed using ClusterProfiler with R programming. (C-E) Bubble heatmaps showing the hierarchical clustering based on Euclidian distance of expression patterns of genes associated to (C) cytokine signaling, (D) chemotaxis, and (E) neutrophil-mediated immunity in COVID-19 and HLH datasets. The color of circles corresponds to log2 fold change (log2FC). Pleiotropic genes belonging to more than one category are bold (Supp. Table S8).

Figure 3: Disease state impacts the correlation between cytokine/chemotaxis and neutrophil-mediated immunity genes. (A and B) Estimated correlations of cytokine signaling/chemotaxis and neutrophil-mediated immunity molecules versus their corresponding first 2 canonical variates (x-CV1 and x-CV2, for cytokine/chemotaxis related genes; γ-CV1 and γ-CV2 for neutrophil-mediated immunity genes in (A) controls and (B) COVID-19 patients. Cytokine/chemotaxis and neutrophil-mediated immunity genes with a correlation of ≥ 0.7 or ≤ - 0.7 are colored in green and blue, respectively, while those with a correlation of < 0.7 or > -0.7 are gray in both groups. (C) Scatter plots with marginal boxplots display the relationship
between variables (genes). Correlation coefficient ($\rho$) and significance level (p-value) for each correlation are shown within each graph.

**Figure 4**: Transcripts stratifying severe COVID-19 from other respiratory diseases are also highly dysregulated in HLH. (A) Schematic overview of study design and patient classification of dataset GSE157103 reported by Overmyer et al.\textsuperscript{58}. (B) Protein-protein interaction (PPI) network among 158 proteins and the 25 genes significant for severe COVID-19_ICU. Node colour denotes Gene Ontology biological process terms. Genes (lower right side half circle) and proteins (left circle and upper right half circle) are denoted by different symbols as explained in the figure legend. Blue circled symbols represent molecules involved in neutrophil-mediated immunity. (C) Principal Component Analysis (PCA) with spectral decomposition shows the stratification of COVID-19_ICU from COVID-19_nonICU and other respiratory diseases (Control_nonICU and Control_ICU). Variables with positive correlation are pointing to the same side of the plot, contrasting with negative correlated variables, which point to opposite sides. Confidence ellipses are shown for each group/category. (D) PCA displaying the stratification of HLH and healthy controls based on the same 25 DEGs as in (B). (E) Correlation matrices of the 25 DEGs (Controls, left matrix; COVID-19_nonICU, middle matrix; and COVID-19_ICU, right matrix). The color scale bar represents the Pearson’s correlation coefficient, containing negative and positive correlations from -1 to 1, respectively. (F) Scatter plots with marginal boxplots display the relationship between the eight genes stratifying severe COVID-19. Correlation coefficient ($\rho$) and significance level (p-value) for each correlation are shown within each graph.

**Figure 5**: Multi-layered transcriptomic analysis associates COVID-19 and HLH common genes with disease severity. (A) Schematic overview of sample cohort and classification of scRNAseq dataset obtained by Schulte-Schrepping et al.\textsuperscript{65} and used for the following analysis. (B) Heatmap showing scRNAseq expression of differentially expressed genes (DEGs) associated with disease severity. Cells and cohorts (controls, mild and severe COVID-19) are indicated by different colors in the legends. (C) Box plots of scRNAseq expression demonstrating that 11 from the 21 genes identified in B are up-regulated when comparing severe and mild COVID-19 patients. (D) Box plots of the 11 transcripts stratifying COVID-19_ICU patients from COVID-19_nonICU patients obtained from the bulk RNAseq dataset from Overmyer et al.\textsuperscript{58}. Significant differences between groups are indicated by asterisks (* $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$). (E) Box plots of microarray data illustrating that the disease severity association of COVID-19 detected by scRNAseq corresponds to the expression differences between HLH patients and controls. Significant differences between groups are indicated by asterisks (* $p \leq 0.05$, ** $p \leq 0.01$ and *** $p \leq 0.001$).

**Figure 6**: Random Forest prediction analysis suggests potential biomarker for severe COVID-19. (A) Receiver operating characteristics (ROC) curve of 11 genes from COVID-19_ICU compared to COVID-19_nonICU patients with an area under the curve (AUC) of 82.4% for both groups. 1=COVID-19_nonICU; 2=COVID-19_ICU. (B) Stable Curve showing number of trees and error rate (out of bag or OOB) with medium of 27,03%. 1=COVID-19_nonICU; 2=COVID-19_ICU.
(C) Variable importance scores plot based on gini_decrease and number(no)_of_nodes for each variable showing which variables are more likely to be essential in the random forest’s prediction. (D) Ranking of the top 10 variables according to mean minimal depth (vertical bar with the mean value in it) calculated using trees. The blue color gradient reveals the min and max minimal depth for each variable. The range of the x-axis is from zero to the maximum number of trees for the feature. (E) Mean minimal depth variable interaction plot showing most frequent occurring interactions between the variables on the left side with light blue color, and least frequent occurring interactions on the right side of the graph with dark blue color. The red horizontal line indicates the smallest mean minimum depth and the black lollipop represent the unconditional mean minimal depth of a variable.
### Table 1: Dataset Information and sample size used for transcriptome analysis

<table>
<thead>
<tr>
<th>Database</th>
<th>Dataset ID</th>
<th>Seq. Method</th>
<th>Sample type</th>
<th>Number of patients</th>
<th>Number of Controls</th>
<th>Original study</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEO</td>
<td>GSE26050</td>
<td>microarray</td>
<td>PBMCs</td>
<td>11</td>
<td>33</td>
<td>Sumegi et al. 2011&lt;sup&gt;107&lt;/sup&gt;</td>
</tr>
<tr>
<td>GEO</td>
<td>GSE152418</td>
<td>bulk-RNA seq</td>
<td>PBMCs</td>
<td>17</td>
<td>17</td>
<td>Arunachalam et al. 2020&lt;sup&gt;108&lt;/sup&gt;</td>
</tr>
<tr>
<td>GEO</td>
<td>GSE152075</td>
<td>bulk-RNA seq</td>
<td>nph swab</td>
<td>430</td>
<td>54</td>
<td>Liebermann et al. 2020&lt;sup&gt;109&lt;/sup&gt;</td>
</tr>
<tr>
<td>GEO</td>
<td>GSE156063</td>
<td>bulk-RNA seq</td>
<td>nph swab</td>
<td>93</td>
<td>141&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Mick et al. 2020&lt;sup&gt;110&lt;/sup&gt;</td>
</tr>
<tr>
<td>GEO</td>
<td>GSE163151</td>
<td>bulk-RNA seq</td>
<td>nph swab</td>
<td>138</td>
<td>11</td>
<td>Ng et al. 2021&lt;sup&gt;61&lt;/sup&gt;</td>
</tr>
<tr>
<td>GEO</td>
<td>GSE152641</td>
<td>bulk-RNA seq</td>
<td>PBL</td>
<td>62</td>
<td>24</td>
<td>Thair et al. 2021&lt;sup&gt;62&lt;/sup&gt;</td>
</tr>
<tr>
<td>GEO</td>
<td>GSE157103</td>
<td>bulk-RNA seq</td>
<td>PBL</td>
<td>100</td>
<td>26&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Overmyer et al. 2020&lt;sup&gt;58&lt;/sup&gt;</td>
</tr>
<tr>
<td>EGA</td>
<td>EGAS00001004571</td>
<td>scRNA seq</td>
<td>PBL/PBMCs</td>
<td>Cohort1 18 Cohort2 17</td>
<td>21</td>
<td>Schulte-Schrepping et al. 2020&lt;sup&gt;65&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>individuals with non-infectious respiratory diseases (NIRD; n=100) and with other respiratory infectious diseases (OIRD; n=41)

<sup>b</sup>individuals negative for SARS-CoV-2 but admitted to ICU (n=16) and nonICU (n=10) units due to respiratory symptoms

nph=nasopharyngeal
METHODS

RESOURCE AVAILABILITY

Data and code availability
This paper analyzes existing, publicly available data. The accession numbers for the datasets are listed in the key resources table.

All original codes used for data analysis have been deposited at github (https://github.com/lschimke/COVID19-and-HLH-paper) and are publicly available as of the date of publication. R packages are listed in the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Data Curation
We searched in public functional genomics data repositories (Gene Expression Omnibus\textsuperscript{111} and Array Express\textsuperscript{112} for human transcriptome data from patients with HLH and COVID-19 published until February 2021. Inclusion criteria for datasets were transcriptome data from minimum n=10 patients without any treatment, availability of gene counts, blood or nasopharyngeal swab samples, and availability of control groups for comparison. This resulted in 7 studies, 6 from COVID-19 patients and one from HLH patients (containing 4 patients with germline mutations and 7 without identified mutation) with transcriptome data generated from different platforms (Table 1). Furthermore, we included the scRNAseq study obtained from the European Genome-phenome Archive\textsuperscript{113} (EGA; EGAS00001004571) of blood leukocytes from two independent cohorts of patients with COVID-19 and healthy controls, respectively\textsuperscript{65}. Combined, we investigated a total of 1253 transcriptome samples.
Differential expression analysis, meta-analysis and visualization of multiple gene expression data sets from microarray and bulk RNAseq

Read counts were transformed (log2 count per million or CPM) and differentially expressed genes (DEGs) between groups were identified through the webtool NetworkAnalyst 3.0\textsuperscript{114} using limma-voom pipeline\textsuperscript{115}. Statistical cut-offs to define DEGs are described below in the section statistical analysis. Shared DEGs among all datasets were displayed using Venn diagram\textsuperscript{116} and Circos Plot\textsuperscript{117} online tools.

Single cell RNAseq analysis

The Seurat Object containing the scRNAseq published by Schulte-Schrepping et al.\textsuperscript{65} and deposited at the EGA (EGAS00001004571) were used for single cell analysis. We followed the Seurat pipeline\textsuperscript{118} as previously described by Stuart et al.\textsuperscript{119} to perform differential expression analysis and data visualization, i.e., UMAP, dotplot, and heatmap construction. Regression for the number of UMIs and scaling were performed as previously described\textsuperscript{65}.

Interactome analysis

For more comprehensive Protein–Protein Interaction (PPI) analyses, we used NAViGaTOR 3.0.14\textsuperscript{120} to visualize genes commonly dysregulated in COVID-19 and HLH datasets, highlighting the biological processes enriched by each gene. Prior to visualization, DEGs were used as input into Integrated Interactions Database (IID version 2021-05; http://ophid.utoronto.ca/iid)\textsuperscript{106,121} to identify direct physical protein interactions. Then, the resultant network was annotated, analysed, and visualized using NAViGaTOR 3.0.14\textsuperscript{120}.

Enrichment analysis and data visualization

We used ClusterProfiler\textsuperscript{122} R package to obtain dot plots of enriched signaling pathways. Elsevier Pathway Collection analysis for selected gene lists (7 genes underlying fHLH/IEI and 11 genes associated with severe COVID-19) was carried out using Enrichr webtool\textsuperscript{123–125}. Set of genes
associated with cytokine/chemotaxis and neutrophil-mediated immunity from each dataset were visualized in bubble-based heat maps applying one minus cosine similarity using Morpheus\textsuperscript{126}. Circular heatmaps were generated using R version 4.0.5 (The R Project for Statistical Computing. https://www.r-project.org/) and R studio Version 1.4.1106 (RStudio. https://www.rstudio.com/) using the circlize R package. Box plots were generated using the R packages ggpubr, lemon, and ggplot2.

**Correlation Analysis**

Principal Component Analysis (PCA) of genes associated with COVID-19 severity (25 transcripts) was performed using the R functions prcomp and princomp, through factoextra package\textsuperscript{127}. Canonical Correlation Analysis (CCA)\textsuperscript{128} of genes associated with cytokines/chemokines and neutrophil-mediated immune responses was performed using the packages CCA and whitening\textsuperscript{128}. In addition, we used the corrgram, psych, and inlmisc R packages to build correlograms. In addition, multilinear regression analysis for combinations of different variables (genes) was performed using the R package ggpubr, ggplot2 and ggExtra.

**Proteome Data Analysis**

We also evaluated the proteomics data obtained from plasma samples of COVID-19 patients previously reported by Overmyer et al.\textsuperscript{58}. Briefly, raw LFQ abundance values were quantified, normalized and log2 transformed as previously described\textsuperscript{58}. Differences in protein expression between COVID-19 ICU and COVID-19 nonICU were calculated as described below in the section _statistical analysis_. Enrichment of proteins significant for COVID-19 ICU was performed using Enrichr webtool\textsuperscript{123–125} and most significant enriched pathways were displayed by dot plot created with R using tidyverse, viridis and ggplot2 packages while Circos Plot of gene-pathway association was built using Circos online tool\textsuperscript{117}. 

(Which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.
Decision-tree classification and machine learning model predictors

We employed random forest model to construct a classifier able to discriminate between COVID-19_nonICU and COVID-19_ICU highlighting the most significant predictors for ICU admission. We trained a Random Forest model using the functionalities of the R package randomForest (version 4.6.14)\textsuperscript{79}. Five thousand trees were used, and the number of variables resampled were equal to three. Follow-up analysis used the Gini decrease, number of nodes, and mean minimum depth as criteria to determine variable importance. Interaction between pair of variables was assessed by minimum depth as criterion. The adequacy of the Random Forest model as a classifier was assessed through out of bags error rate and ROC curve. For cross-validation, we split the dataset in training and testing samples, using 75% of the observations for training and 25% for testing.

STATISTICAL ANALYSIS

Differentially expressed genes and proteins

To determine DEGs of each dataset we applied the statistical cut-offs of log2 fold-change > 1 (up-regulated), log2 fold-change < -1 (down-regulated), and adjusted p-value < 0.05. We used the Fisher’s method to combine p values from multiple studies for information integration\textsuperscript{114}. Differences in protein expression between COVID-19_ICU and COVID-19_nonICU was calculated using the nonparametric MANOVA (multivariate analysis of variance) test\textsuperscript{129} followed by analysis of nonparametric Inference for Multivariate Data\textsuperscript{130} using the R packages npmv, nparcomp, and ggplot2.
SUPPLEMENTAL INFORMATION

All supplemental figures, titles, and legends are provided in separate document file.

Supplemental Tables S1-S14 are provided in a separate document file:
Table S1: Up- and Down-regulated DEGs of all datasets
Table S2: List of common DEGs
Table S3: Genelist Overlaps
Table S4: Biological Process (BP) enriched terms
Table S5: Cellular components (CC) enriched terms
Table S6: Cytokine and Chemotaxis genes
Table S7: Neutrophil-mediated immunity genes
Table S8: Pleiotropic genes
Table S9: DEGs between COVID-19_ICU and COVID-19_nonICU
Table S10: DEGs in additional datasets
Table S11: 25 Genes in proteomic/protein expression datasets
Table S12: Significant plasma proteins comparing COVID-19_ICU versus COVID-19_nonICU patients
Table S13: Enriched BP GO terms of 158 significant plasma proteins
Table S14: Enriched Elsevier Pathway collections terms of 11 genes significant for severe COVID-19
REFERENCES


78. Asad, S. et al. Proteomics-Informed Identification of Luminal Targets For In Situ Diagnosis...


89. Li, Y. et al. SARS-CoV-2 early infection signature identified potential key infection mechanisms and drug targets. BMC Genomics 22, (2021).


113. Lappalainen, I. et al. The European Genome-phenome Archive of human data consented


**Graphic abstract**

**DATASETS**
- fHLH
- COVID19
- Controls

- GSE26050
- GSE152418, GSE157103
- GSE152075, GSE150603
- GSE163151, GSE152641
- EGA50000104571

**MULTI-LAYERED TRANSCRIPTOME ANALYSIS**
- scRNAseq
- bulk RNAseq
- Microarray

**TRANSCRIPTIONAL OVERLAP**
- Molecular basis for phenotypic convergence
  - 239 common DEGs and signaling pathways
    - chemokinesis/cytokine and neutrophil hyperactivation

**DISEASE SEVERITY**
- Transcriptional differentiation between COVID-19 ICU and nonICU patients
- Multi-layered transcriptomic and proteomic network of molecules significant for severe COVID-19
- Machine learning model

**1253 Individuals**
- 693 patients
- 360 controls

**25481 DEGs**
- PBMCs
- PBLs
- Swabs

**239 common DEGs**
- 237 up-regulated
- 2 down-regulated

**25 GENES**
- Stratifying COVID-19 ICU from nonICU patients
- Significant for severe COVID-19 across different studies
Phenotype convergence of COVID-19 and HLH

Up-regulated
Down-regulated

289 492
3107
613 213
4537
3899
3727
735
978
376
1118
716
423
355
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

A: PBMCs_GSE152418
B: PBL_GSE157103
C: swab_GSE152075
D: swab2 NIRD_GSE156063
E: swab2 OIRD_GSE156063
F: scRNA cohort1_EGAS00001004571
G: scRNA cohort2_EGAS00001004571
H: HLH_GSE26050

Node color: GO Biological Process

Fig 1

Fig 2
Fig 2

Cytokine signaling
- neutrophil activation
- neutrophil-mediated immunity
- neutrophil degranulation
- neutrophil activation involved in immune response
- response to lipopolysaccharide
- leukocyte migration
- cellular response to lipopolysaccharide
- cellular response to molecule of bacterial origin
- cellular response to biotic stimulus
- positive regulation of cytokine production
- cell chemotaxis
- humoral immune response
- response to interleukin-1
- cellular response to interleukin-1
- leukocyte chemotaxis
- myeloid leukocyte migration
- granulocyte migration
- neutrophil migration
- antimicrobial humoral response
Fig 4

A

- Mild
  - Respiratory symptoms
  - Medical floor unit
- Severe
  - Respiratory symptoms
  - Intensive care unit
  - Severity Index: APACHE II, SOFA

B

- CONTROL_nonICU (n=10)
  - SARS-CoV-2 negative
- CONTROL_ICU (n=15)
- COVID-19_nonICU (n=50)
- COVID-19_ICU (n=50)

C

- DEFA4
- LCN2

D

- DEFA4

E

- CONTROLS
- COVID-19_nonICU
- COVID-19_ICU

F

- ELANE
- AZU1
- CTSG
- CEACAM8
- MMP8
**Fig 5**

E. Control Severe Mild

bulk RNAseq (COVID-19 cohort; GSE157103, Overmyer et al. 2020)

---

C. scRNAseq (COVID-19 cohort 1 and 2; EGAS00001004571, Schulte-Schrepping et al. 2020)

D. bulk RNAseq (COVID-19 cohort; GSE157103, Overmyer et al. 2020)

E. Microarray (HLH cohort; GSE26050, Sumegi et al. 2011)
**KEY RESOURCES TABLE**

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deposited data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>microarray data of HLH patients and healthy controls</td>
<td>Sumegi et al. 2011</td>
<td>GEO: GSE26050</td>
</tr>
<tr>
<td>bulk-RNA seq data of COVID-19 patients and controls</td>
<td>Arunachalam et al. 2020</td>
<td>GEO: GSE152418</td>
</tr>
<tr>
<td>bulk-RNA seq data of COVID-19 patients and controls</td>
<td>Liebermann et al. 2020</td>
<td>GEO: GSE152075</td>
</tr>
<tr>
<td>bulk-RNA seq data of COVID-19 patients and patients with other respiratory diseases</td>
<td>Mick et al. 2020</td>
<td>GEO: GSE156063</td>
</tr>
<tr>
<td>bulk-RNA seq data of COVID-19 patients and controls</td>
<td>Ng et al. 2021</td>
<td>GEO: GSE163151</td>
</tr>
<tr>
<td>bulk-RNA seq data of COVID-19 patients and controls</td>
<td>Thair et al. 2021</td>
<td>GEO: GSE152641</td>
</tr>
<tr>
<td>bulk-RNA seq data of COVID-19 patients and patients with other respiratory symptoms</td>
<td>Overmyer et al. 2020</td>
<td>GEO: GSE157103</td>
</tr>
<tr>
<td>scRNA seq data of COVID-19 patients and controls</td>
<td>Schulte-Schrepping et al. 2020</td>
<td>EGA: EGAS00001004571</td>
</tr>
<tr>
<td>Software and algorithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R version 4.0.5</td>
<td>R Core Team (2020)</td>
<td><a href="https://www.r-project.org/">https://www.r-project.org/</a></td>
</tr>
<tr>
<td>circlize R package</td>
<td>Zuguang Gu et al. 2020</td>
<td><a href="https://github.com/jokergoo/circlize">https://github.com/jokergoo/circlize</a></td>
</tr>
<tr>
<td>Package</td>
<td>Author(s)</td>
<td>Year</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>ggplot2 R package</td>
<td>Wickham H, 2016</td>
<td></td>
</tr>
<tr>
<td>corrgram R package</td>
<td>Kevin Wright, 2021</td>
<td></td>
</tr>
<tr>
<td>psych R package</td>
<td>William Revelle, 2021</td>
<td></td>
</tr>
<tr>
<td>tydiverse R package</td>
<td>Hadley Wickham, 2021</td>
<td></td>
</tr>
<tr>
<td>nonparametric MANOVA: nparcomp R package</td>
<td>Konietschke, F. et al., 2015</td>
<td></td>
</tr>
</tbody>
</table>