Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients

Anna C. Aschenbrenner[1,2,*], Maria Mouktaroudi[3,*], Benjamin Krämer[4,*], Nikolaos Antonakos[3,*], Marie Oestreich[1,*], Konstantina Gkizeli[3,*], Melanie Nuesch-Germano[1,*], Maria Saridaki[3,*], Lorenzo Bonaguro[1,*], Nico Reusch[1,*], Kevin Baßler[1,*], Sarandia Doulou[3,*], Rainer Knoll[1,*], Tal Pecht[1,*], Theodore S. Kapellos[1,*], Nikoletta Rovina[5,*], Charlotte Kröger[1,*], Miriam Herbert[1,*], Lisa Holsten[1,*], Arik Horne[1,*], Ioanna D. Gemünd[1,*], Shobhit Agrawal[1,*], Kilian Dahm[1,*], Martina van Uelft[1,*], Anna Drews[6,*], Lena Lenkeit[1,*], Niklas Bruse[7,*], Jelle Gerretsen[7], Jannik Gierlich[1], Matthias Becker[6], Kristian Händler[6], Michael Kraut[6], Heidi Theis[6], Simachew Mengiste[1], Elena De Domenico[6], Jonas Schulte-Schrepping[1], Lea Seep[1], Jan Raabe[4], Christoph Hoffmeister[4], Michael ToVinh[4], Verena Keitel[8], Gereon Rieke[4], Valentina Talevi[9], N. Ahmad Aziz[9,10], Peter Pickkers[7], Frank van de Veerdonk[2], Mihai G. Netea[2,11,#], Joachim L. Schultze[1,6,#], Matthijs Kox[7,#], Monique M.B. Breteler[8,12,#], Jacob Nattermann[4,#], Antonia Koutsoukou[5,#], Evangelos J. Giamarellos-Bourboulis[3,#], Thomas Ulas[1,6,#§], German COVID-19 Omics Initiative (DeCOI)

Affiliations

[1] Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, 53115 Bonn, Germany

[2] Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen 6525, The Netherlands

[3] 4th Department of Internal Medicine, National and Kapodistrian University of Athens, Medical School, 124 62 Athens, Greece

[4] Department I of Internal Medicine, University Hospital of Bonn (UKB), Venusberg-Campus-1, 53127, Bonn, Germany

[5] 1st Department of Pulmonary Medicine and Intensive Care Unit, National and Kapodistrian University of Athens, Medical School, 115 27, Greece

[6] Platform for Single Cell Genomics and Epigenomics at the German Center for Neurodegenerative Diseases and the University of Bonn, 53127 Bonn, Germany

[7] Department of Intensive Care Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen 6500 HB, The Netherlands

[8] Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[9] Population Health Sciences, German Center for Neurodegenerative diseases (DZNE), Bonn, 53127 Bonn, Germany

[10] Department of Neurology, Faculty of Medicine, University of Bonn, 53127 Bonn, Germany

[11] Immunology & Metabolism, Life and Medical Sciences (LIMES) Institute, University of Bonn, Bonn 53115, Germany

[12] Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Bonn, Germany

[13] German Center for Infection Research (DZIF), Germany

[*] shared first authorship

[#] shared last authorship

[§] Corresponding author: t.ulas@uni-bonn.de

SUMMARY

The SARS-CoV-2 pandemic is currently leading to increasing numbers of COVID-19 patients all over the world. Clinical presentations range from asymptomatic, mild respiratory tract infection, to severe cases with acute respiratory distress syndrome, respiratory failure, and death. Reports on a dysregulated immune system in the severe cases calls for a better characterization and understanding of the changes in the immune system. Here, we profiled whole blood transcriptomes of 39 COVID-19 patients and 10 control donors enabling a data-driven stratification based on molecular phenotype. Neutrophil activation-associated signatures were prominently enriched in severe patient groups, which was corroborated in whole blood transcriptomes from an independent second cohort of 30 as well as in granulocyte samples from a third cohort of 11 COVID-19 patients. Comparison of COVID-19 blood transcriptomes with those of a collection of over 2,600 samples derived from 11 different viral infections, inflammatory diseases and independent control samples revealed highly specific transcriptome signatures for COVID-19. Further, stratified transcriptomes predicted patient subgroup-specific drug candidates targeting the dysregulated systemic immune response of the host.

KEYWORDS

COVID-19; blood transcriptomics; transcriptome; co-expression analysis; stratification; molecular disease phenotypes; granulocytes; neutrophils; drug repurposing

64 # INTRODUCTION

65 Pandemic spread of the recently emerged coronavirus, severe acute respiratory syndrome-coronavirus 2
66 (SARS-CoV-2), has resulted in over 9.2 million confirmed infected individuals and over 470,000 deaths
67 worldwide (WHO, covid19.who.int, as of June 24th, 2020) from the resulting severe respiratory illness, called
68 coronavirus disease 2019 (COVID-19) (*1–3*). Based on clinical observations, it has become clear that there
69 is great variety in disease manifestation, ranging from asymptomatic cases, to flu-like symptoms, to severe
70 cases needing mechanical ventilation, to those who do not survive (*4–8*). Increasing evidence suggests
71 that the immune system plays a pivotal role in determining the severity of the disease course and it has
72 been suggested that different molecular phenotypes might be responsible for the heterogenous outcome
73 of COVID-19 (*9–11*). Identifying these molecular phenotypes might not only be important for a better
74 understanding of the pathophysiology of the disease, but also to better define patient subgroups that are
75 more likely to benefit from specific therapies (*12–17*). Indeed, while vaccines are still under development,
76 finding an effective and patient-tailored therapeutic management for COVID-19 patients including targeting
77 derailed immune mechanisms (*18*, *19*) is key to mitigate the clinical burden as well as to prevent further
78 disease fatalities (*15*, *16*).

79 The analysis of peripheral blood-derived immune parameters in inflammatory and infectious diseases either
80 by classical testing, including flow cytometry and serum protein measurements, or omics technologies,
81 including transcriptomics, has been proven very valuable in the past (*20–28*). In COVID-19 patients,
82 monitoring peripheral blood as a proxy for the ongoing changes within the circulating cells of the immune
83 system, has revealed lymphopenia to correlate with disease severity (*29*). Similarly to SARS-CoV and
84 MERS-CoV infections, hyperinflammation due to excessive release of proinflammatory cytokines is often
85 observed in severe COVID-19 patients as increased serum IL-6 levels correlate with respiratory failure and
86 adverse clinical outcomes (*9*, *30*, *31*).

87 While one can envision mild and/or early cases to benefit from antiviral drug treatments currently under
88 clinical investigation, more severe cases may benefit from treatment to mitigate the excessive systemic
89 immune reactions resulting in progressing pneumonia and even respiratory failure associated with severe
90 COVID-19 (*4–9*). The detrimental role of the systemic inflammation in the late phase of the disease has
91 become clear, as the cytokine storm has been associated with disease morbidity (*6*, *9*, *30–33*). Thus, a
92 better understanding of the dysregulation of the host response to the infection leading to immunopathology
93 is urgently needed to dissect and comprehend the immune parameters accompanying the heterogeneous
94 disease severity seen upon SARS-CoV-2 infection.

95 Based on previous experience with other infectious diseases (*20–26*), we hypothesized that whole blood
96 transcriptomes should allow us to 1) determine immune cellular characteristics and functions in COVID-19
97 patients, 2) reveal heterogeneous molecular phenotypes of patients with similar clinical presentation, 3)

define commonalities and differences of COVID-19 in comparison to other inflammatory conditions and 4) predict potential drug repurposing that might counteract observed immune dysregulations.

Here, by using blood transcriptomes, we provide evidence for molecular subtypes within the immune response of COVID-19 patients beyond distinguishing mild and severe cases only. In addition, molecular changes in blood of severely affected patients are strikingly associated with changes in the granulocyte compartment. Furthermore, blood transcriptomes of molecular subtypes of COVID-19 patients seem to be unique in comparison to more than 2,600 samples derived from other infections, inflammatory conditions and controls. Finally, by reverse drug target prediction using patients' blood transcriptomes revealed known as well as additional new potential targets for further evaluation. Our data might also serve as a starting point for a large-scale assembly of molecular data collected during currently ongoing and future therapy trials for COVID-19 patients based on whole blood transcriptomes.

RESULTS

Whole blood transcriptomes reveal diversity of COVID-19 patients not explained by disease severity

To investigate the host immune response of COVID-19 patients in a systematic approach, whole blood transcriptomes were analyzed from 39 patients and 10 control donors recruited at the same hospital by RNA-sequencing (RNA-seq, **Fig. 1A**). Two-dimensional data representation using principal component analysis (PCA) showed separation of COVID-19 and control samples (**Fig. 1B**). Differential expression analysis identified 2,289 upregulated and 912 downregulated genes comparing COVID-19 and control samples (FC>|2|, padj<0.05 / **Fig. 1C**). Upregulated genes showed greater fold changes than the downregulated genes (**Fig. 1D**). Of note, *CD177*, markedly expressed in neutrophils (*34*, *35*), was the most prominently upregulated gene with the lowest p-value. Heightened expression was further found for several granulocyte- and monocyte-associated molecules, such as Eosinophil-derived neurotoxin (*RNASE2*), Haptoglobin (*HP*), Neutrophil elastase (*ELANE*), Olfactomedin 4 (*OLFM4*), Myeloperoxidase (*MPO*), Resistin (*RETN*), matrix metalloproteinases (*MMP8*, *MMP9*), and alarmins (*S100A8*, *S100A9*, *S100A12*), as well as for cell cycle progression-associated genes (*G0S2*, *CDC6*, *CDC25A*), type I interferon (IFN)-induced genes (*IFI27*, *IFITM3*, CD169 (*SIGLEC1*)), but also genes with immunosuppressive functions (*IL10*, *SOCS3*, Arginase (*ARG1*)). Downregulated genes included many lymphocyte-associated factors, such as *NELL2*, *RORC*, *KLRB1*, TCF1 (*TCF7*), Calcipressin-3 (*RCAN3*), *BACH2*, or *LEF1* (**Fig. 1D**, **Table S1**). Functional analysis of the differentially expressed genes (DEGs) by gene ontology enrichment analysis (GOEA) revealed granulocyte and complement activation-associated terms enriched in the upregulated DEGs and lymphocyte differentiation and T cell activation for the downregulated DEGs (**Fig. 1E**).

131 Interestingly, the T cell activation-associated genes accounting for the enrichment of this term for the
132 upregulated DEGs included *IL10* and *CD274* (PD-L1) pointing at suppressive T cell functionality (**Table**
133 **S1**).

134 In the current cohort, 51% of COVID-19 patients required intubation (**Table S2**). Given the heterogeneous
135 nature of clinical manifestation of COVID-19, we sought to stratify the transcriptomic profiles by disease
136 severity based on intubation status. Indeed, samples from patients with mild disease (requiring no
137 intubation) clustered more closely to the control samples, while those of severe cases scattered away in
138 the PCA (**Fig. 1F**). Consequently, there was a greater number of DEGs in blood samples from severe
139 COVID-19 patients than in mild patients when compared to controls (**Fig. 1G**). Many of the DEGs found in
140 the COVID-19 vs control comparison (**Fig. 1D**) were also found when separating the COVID-19 samples
141 by severity (**Fig S1A,B**). Both, severe and mild COVID-19 in comparison to controls shared neutrophil-
142 specific *CD177* and *HP* expression among the most upregulated DEGs, as well as lymphocyte-associated
143 genes such as ABLIM1, NELL2, RCAN3, RORC, KLRB1, among the downregulated genes (**Fig. S1A**,**B**).
144 GOEA reflected these findings (**Fig. S1C**). Although all samples from COVID-19 patients showed functional
145 enrichment for granulocyte/neutrophil activation-associated terms in general, direct comparison of severe
146 and mild COVID-19 patients revealed this to be a heightened characteristic of the immunoprofiles in severe
147 COVID-19 (**Fig. S1C**). Upregulated DEGs in the severe vs. mild sample comparison included *CD177*,
148 Haptoglobin (*HP*), Neutrophil elastase (*ELANE*), Olfactomedin 4 (*OLFM4*), Myeloperoxidase (*MPO*),
149 Resistin (*RETN*), matrix metalloproteinase *MMP8*, and alarmins (*S100A8*, *S100A12*). Whereas the type I
150 IFN-response genes, such as *IFI27* or *IFITM3*, were not differentially regulated in severe vs. mild samples,
151 expression of immunosuppression-associated factors was more pronounced in severe COVID-19 patients
152 (*IL10*, *SOCS3*, Arginase (*ARG1*)) (**Fig. 1H, Table S1**). Moreover, blood transcriptomes from severe cases
153 showed decreased expression of lymphocyte-associated genes, such as the T cell receptor chains (*TRAC*,
154 *TRBC1*), CD3 zeta chain (*CD247*), *CD4*, *CD2*, *IL2RB*, TBET (*TBX21*), *IL7R*, as well as monocyte-
155 associated genes, such as MHC class II molecules (*HLA-DPA1*, *HLA-DRB5*), fractalkine receptor
156 (*CX3CR1*), Macrophage scavenger receptor (*MSR1*), or *CCL2* (**Fig. 1H, Table S1**). Differences in gene
157 expression were not restricted to granulocyte and T cell functions only: assessing the changes in defined
158 gene groups, e.g. transcription factors, epigenetic regulators, surface or secreted molecules, we observed
159 many significant changes in genes that are not restricted to granulocytes or T cells, clearly indicating that
160 other cell types are also transcriptionally altered in COVID-19 patients (**Fig. S1D**).

161 Distribution of the COVID-19 samples in the PCA revealed heterogeneity in the transcriptomic profiles (**Fig.**
162 **1F**), which might be due to clinical heterogeneity (**Table S2**). In order to investigate this further, the top 25%
163 of the most variable expressed genes were visualized in a heat map and samples sorted by unbiased
164 hierarchical clustering based on their transcriptomic profiles, which resulted in more than three clusters
165 suggesting higher transcriptional heterogeneity as explained by mild and severe COVID-19 cases vs control

166 (**Fig. 1I**). Strikingly, neither disease, disease severity, nor the inclusion of outcome or immune classification
167 (*31*), sufficiently explained the structure in the data. In order to get a better clinical understanding of the
168 transcriptional data, we included further clinical parameters and grouped the COVID-19 patients
169 accordingly (**Fig. 1I**). We therefore performed agglomerative hierarchical clustering using the clinical
170 parameters that contributed most to the transcriptional differences observed across the first principal
171 component of the dataset (r-adjusted square ≥0.1, **Fig. S1E**). The COVID-19 patients were clustered into
172 five clinical groups, which was the optimal number of clusters at which the intra-group variance was low
173 and the 'clusters distance' remained high (**Fig. S1F,G**). However, comparison of this clinical parameter-
174 based grouping of the COVID-19 patients did not match the transcriptional variability observed in the data
175 either (**Fig. 1I**), arguing that additional molecular parameters must exist that better define the blood
176 transcriptome structure and thereby more accurately dissect heterogeneity of the clinical manifestation of
177 COVID-19.

178

179 ## Co-expression analysis discloses COVID-19 subgroups with distinct molecular signatures

180 Classical approaches to analyze the transcriptome data by using differential gene expression analysis
181 based on sample groups defined by a selection of clinical parameters precluded dissection of the
182 heterogeneity of the host immune response towards SARS-CoV-2 infection, which is evident in the high-
183 parameter space of the transcriptome (**Fig. 1**). Co-expression analysis on the other hand identifies similarly
184 regulated genes across samples, groups these genes into modules, which can then be explored for each
185 patient sample individually or for entire patient groups. Applying such an approach using our established
186 CoCena² pipeline [https://github.com/Ulas-lab/CoCena2] (**Fig. 2A**) for all 49 samples (39 COVID-19, 10
187 control) independent of their clinical annotation disclosed 10 co-expression modules, designated by color
188 indianred to darkgrey, across a total of 6,085 genes included in the analysis (**Fig. S2A**). Hierarchical
189 clustering of the samples based on their group fold changes (GFCs) for each module revealed a data-driven
190 patient stratification assorting the samples into six groups (**Fig. S2B**), which were subsequently used in all
191 following analyses: five different COVID-19 sample-containing groups, which only partially grouped by
192 disease severity and illustrated heterogeneity of the immune response in COVID-19 patients, plus one
193 group containing all control as well as four COVID-19 samples (**Fig. 2B**+**S2C**). Overlaying this information
194 onto the original PCA reflected structured sample stratification as the newly defined groups clustered
195 together (**Fig. S2D**). GFC analysis of the newly generated groups revealed group-specific enrichment of
196 co-expressed gene modules (**Fig. 2C**). GOEA on each of the modules identified associated gene signatures
197 displaying distinct functional characteristics, which distinguish the different sample groups G1-G6 (**Fig.**
198 **2D+S3, Table S3**). For example, 'inflammatory response' was enriched in modules maroon, lightgreen,
199 pink, and darkgrey, all characteristic for sample groups G1 and G2 to different extents, indicating these to
200 possibly undergoing a more vigorous inflammatory immune reaction (**Fig. 2C+D**). Of note, G1 and G2

201 harbour a great fraction of samples from patients with severe COVID-19 (**Fig. 2B**). Only a slight increase
202 in the inflammation-associated module maroon, an increase in expression in the genes of darkorange
203 (enriched in oxidative phosphorylation, mTORC1 signaling and cell cycle-associated genes), as well as a
204 loss of expression in the gold module (connected to estrogen response genes and IL2 signaling) was
205 indicative of the G4 sample group. G6, encompassing all control samples, was not associated with any
206 modules connected to inflammatory processes, but showed higher expression of indianred, steelblue and
207 gold, all functionally enriched basic cellular and metabolic processes. Extended analysis of the lightgreen
208 module, containing 987 genes, revealed a prominent enrichment of granulocyte/neutrophil activation-
209 related signatures (**Fig. 2E, Table S3**). To further explore this neutrophil activation signature association,
210 we investigated possible co-expression patterns of long non-coding RNAs (lncRNA) that were reported as
211 regulators of granulocyte function (*36*). *CYTOR* (also known as *Morrbid*) is a lncRNA that mediates survival
212 of neutrophils, eosinophils, and classical monocytes in response to pro-survival cytokines (*36*), and
213 interacts with the protein-coding RNAs for the catalytic PI3K isoform Phosphatidylinositol-4,5-bisphosphate
214 3-kinase catalytic subunit beta (*PIK3CB*) and the filament Vimentin (*VIM*) (*37*). Interestingly, expression of
215 *CYTOR* was significantly increased in severe COVID-19 patient group G1 ($p<0.001$) and correlated with
216 both *PIK3CB* ($r= 0.53$, $p<0.001$) and *VIM* ($r= 0.55$, $p<0.001$) (**Fig. 2F**).

217 Next, we asked whether the enrichment for neutrophil activation-associated signatures in G1 and G2 is
218 attributed to an increased relative number of granulocytes within the whole blood sample. Deconvolution of
219 the expression values using linear support vector regression (*38*) showed increased relative percentages
220 of neutrophils especially in G1 and G2 (**Fig. S2E**). G5, on the other hand, clearly displayed an increased
221 percentage of monocytes. At the same time, lymphocyte enrichment was found to be reduced in the COVID-
222 19 sample groups, most prominently in G1 and G2 (**Fig. S2E**). The linear deconvolution results were then
223 validated by flow cytometry. Blood composition of COVID-19 donors confirmed an increased number of
224 neutrophils and a decreased number of lymphocytes especially in G1 and G2 (**Fig. S2F**). As a result, the
225 neutrophil-lymphocyte ratio (NLR), a clinical marker proposed for disease severity as it has been associated
226 with an increased systemic inflammation (*39*, *40*), was markedly elevated in G1 and G2 compared to the
227 control sample-containing G6, both in the computationally deconvoluted results (**Fig. 2G**) as well as
228 measured by flow cytometry (**Fig. 2H**). Interestingly, in context of the observation that men more often
229 progress to severe COVID-19 than women (*41*), G1 encompasses samples from solely male patients (**Fig.
230 S2C**). Analysis of the top 20 differentially expressed transcription factors, epigenetic regulators, surface or
231 secreted proteins for the six sample groups confirmed an increased inflammatory state, again most
232 remarkably for G1 and G2, evident from the transcription factors of the STAT family, *STAT1*, *STAT3*,
233 *STAT5B* and *STAT6*, surface marker *CSF3R* (G-CSF) or *FCGR3B* (CD16b), the secreted factors *GRN* or
234 *IL1B*, or the epigenetic regulator *PADI4* (PAD4) (**Fig. S2H**).

235 We confirmed our findings of distinct molecular phenotypes in the blood of COVID-19 patients in a second
236 independent cohort. Thirty patients, severely affected by SARS-CoV-2 infection, were sampled upon

237 admission to the ICU. We stratified the obtained blood transcriptomes based on the module signatures from
238 the co-expression analysis (**Fig. 2C**). The samples of the second cohort were filtered for the genes present
239 in the COVID-19 co-expression network, group fold changes were calculated across all patients individually,
240 and sample groups G1-G6 assigned according to their combinatorial module expression (**Fig. S4A**).
241 Controls from the first cohort were included for comparison. Interestingly, in these ICU patients, we noted
242 the transcriptome profiles from the second cohort to show greatest similarity to G1 and G2, which is in line
243 with their severe phenotypes and our findings from the first cohort. Hierarchical clustering of the samples
244 based on their group fold changes for each module stratified the samples of the second cohort into four
245 groups (**Fig. S4B**). The control samples from the first cohort built one separate group, which we designated
246 again as G6. To allow for group-specific comparison to the stratification within the first cohort (**Fig. 2C**), we
247 calculated the mean GFCs of the four groups identified in the second cohort (**Fig. S4C**). Second cohort
248 samples of the first group showed enrichment in modules lightgreen, pink and darkgrey and were thus
249 assigned most similar to G1; the third group of the new samples showed enrichment in modules maroon
250 and darkorange, most similar to G2; and the remaining samples were stratified into an intermediate group
251 exhibiting stronger expression of genes from the darkorange as well as pink module indicating
252 characteristics of both G1 and G2 (**Fig. S4C**).

253 Collectively, co-expression analysis (CoCena[2]) in whole blood transcriptomes reveals at least five
254 molecular phenotypes of the host's immune response in COVID-19 patients with at least two different
255 groups in clinically described severe COVID-19 patients. The two molecularly defined groups G1 and G2
256 are transcriptionally characterized by a pronounced neutrophilic signature, at the same time distinct in other
257 cellular characteristics. Such molecular classification might serve as a basis for identifying clinical
258 surrogates for patient stratification. Since whole blood transcriptomics captures functional changes in the
259 host's peripheral immune response across all cell types, we next sought a more detailed investigation of
260 the granulocyte compartment within the framework of the newly identified subgroups.

261

262 Granulocytes from severe COVID-19 patients show a simultaneous increase in inflammatory and 263 suppressive signatures

264 To investigate whether the activation signatures seen in whole blood of COVID-19 patients are not only due
265 to disease-associated increase of the neutrophil population, granulocytes were sequenced and
266 transcriptomes were analyzed from 11 longitudinally sampled patients (4 mild, 7 severe), resulting in 14
267 mild and 45 severe COVID-19 samples (**Fig. 3A**). Evaluation of the relative cell type composition within
268 each sample using linear deconvolution predicted the samples to mainly consist of neutrophils, with
269 comparable fractions of 78.8% on average (**Fig S5A**). Exploratory analysis by PCA showed a separation
270 between mild and severe COVID-19 patients' granulocyte samples, especially for the day 1-14 groups (**Fig.**

271 **3B**). Differential expression analysis identified 1,496 upregulated and 1,440 downregulated genes
272 comparing severe and mild samples from day 1-14 after first symptoms, while comparison at a late disease
273 stage showed less differences on gene level (380 up-, 307 downregulated genes / FC>|2|, padj<0.05 / **Fig.**
274 **3C, Table S4**). Whole blood transcriptome analysis showed enrichment of neutrophil activation-associated
275 signatures (**Fig. 2**). Excluding the bias of alterations in neutrophil population size across conditions, gene
276 set enrichment analysis on granulocyte samples now uncovered that differentially expressed genes
277 between severe and mild COVID-19 patients are indeed characterized by an increase in granulocyte
278 activation-associated factors (**Fig. S5B**). *CD177* is part of the granulocyte activation gene set and was
279 indeed markedly increased in severe (day 1-14) compared to mild (day 1-14) COVID-19 samples (**Fig. 3D**).
280 Also, the alarmin *S100A6* exhibited heightened expression in granulocytes from severe COVID-19 patients
281 (**Fig. 3D**).

282 Next, we used the CoCena[2] modules from the whole blood analysis (**Fig. 2C**) to identify modules that are
283 actually driven by alterations in neutrophil activation instead of a mere increase in the neutrophil population.
284 The genes from the modules were filtered by the union of the upregulated genes between severe and mild
285 COVID-19 patients (either comparing at day 1-14 or 15-28 / **Fig. S5C**). After filtering, the number of DEG
286 between granulocyte samples from mild and severe COVID-19 exceeded 100 genes per module in three
287 of the modules. Among those modules, the genes identified in whole blood transcriptomes within the
288 lightgreen module showed the highest overlap of 45%, maroon of 29% and pink of 7% with genes
289 upregulated in granulocytes of severe COVID-19 patients. We then investigated the expression pattern of
290 those modules for each individual patient in a longitudinal fashion (**Fig. 3E**). In concordance with the whole
291 blood CoCena[2] results (**Fig. 2C**), modules lightgreen, maroon and pink showed a continuously elevated
292 mean expression in the severe compared to the mild COVID-19 patients, indicated by horizontal lines
293 showing the mean expression of the respective modules calculated only for the mild patients (**Fig. 3E**). This
294 effect was observed irrespective of donor and of days after symptoms onset.

295 Recently, heterogeneity of neutrophils with distinct subsets associated with disease severity and phase
296 was revealed by single cell RNA-seq analysis in blood of COVID-19 patients (*42*). Enrichment of the three
297 signatures that related to severe COVID-19, in our granulocyte samples demonstrated that the findings
298 obtained in the single-cell study were also discernible in bulk data and the results in accordance to the
299 reported phenotypes: premature/immature, severe inflammatory, as well as severe suppressive subset
300 marker genes were markedly enriched in granulocytes from severe COVID-19 patients in the present study
301 (**Fig. S5D**). Further analysis of this observation on the gene level displayed the heightened expression of
302 pre-/immature neutrophil-associated markers in severe COVID-19 patients' granulocytes, such as CD15
303 (*FUT4*), metalloproteinase *MMP8*, alarmins (*S100A8/9*), NET formation-involved *PADI4*, or *NLRC4*, for
304 which activating mutations have been reported to overtly trigger the inflammasome and thereby increase
305 the risk to develop autoinflammatory syndrome (*43*, *44*) (**Fig. 3F**). Marker genes attributed to the "mild

306 mature activated" neutrophil subset (*42*), such as *ITGA4*, or *SLC38A1*, were indeed elevated as well in the
307 mild COVID-19 patients' granulocytes of this study. In line with the single cell study, signs of an interferon
308 response were observed irrespective of disease severity (*IFIT1*, *IFIT3*, *ISG15*), while only severe COVID-
309 19 patients' granulocytes featured expression of genes with suppressive functionality, such as *ARG1* or
310 PD-L1 (*CD274*) (**Fig. 3F**),

311 We next stratified the granulocyte samples based on the module signatures from the whole blood analysis.
312 The granulocyte samples were filtered for the genes present in the COVID-19 co-expression network (**Fig.**
313 **2C**) and the group fold changes were calculated across all patients individually, sample groups G1-G6 were
314 assigned according to their combinatorial module expression (**Fig. 2C+3G**). For example, samples
315 attributed to G1 showed high enrichment scores in modules lightgreen, darkgreen and pink, whereas those
316 assigned as G2 additionally expressed the maroon module. Samples with the indianred/darkorange
317 combination were designated as G4. Assessment of the combinatorial enrichment scores for the different
318 modules did not lead to a corresponding sample group for all longitudinal samples from patient 1, hence it
319 was assigned as G7. Re-analysis of *CD177*, *NLRC4*, *ARG1*, and PD-L1 (*CD274*) as a function of the
320 assigned sample groups (**Fig. 3B-D**), showed increased expression in G1 and G2 in relation to the other
321 groups (**Fig. 3H+S5E**). Interestingly, the stratified patient groups in the whole blood data also depicted
322 increased expression in G1 and G2 in comparison to the control-containing G6 (**Fig. 3I+S5F**).

323 Analysis of granulocyte samples from COVID-19 patients proved that, in addition to the relative increase in
324 neutrophils in severe COVID-19 cases, there are indeed alterations in the transcriptional program of these
325 cells themselves. We found enrichment of signatures typical of pre-/immature neutrophils and evidence of
326 simultaneous inflammatory and suppressive features, arguing for a dysregulation in the peripheral
327 granulocyte compartment. Importantly, transferring these findings back to the whole blood analysis showed
328 that the granulocyte phenotypes were still observable within the whole blood transcriptomes.

329

330 Integration with signatures from other diseases reveals COVID-19-specific characteristics

331

332 Putting COVID-19 into context of other known diseases, we compiled whole blood transcriptomes from 11
333 further diseases, including several viral and bacterial infections as well as immune-related disorders into
334 one large dataset encompassing a total of 2,817 samples including the 39 COVID-19 samples from this
335 study (**Fig. 4A**, **S6A, Table S5**). All in all, the dataset contains three other viral infection studies
336 (Chikungunya (*26*), HIV (*23*) and Zika (*45*), n=466), seven bacterial infection studies (tuberculosis (*20–23*,
337 *46*), bacterial sepsis and systemic inflammatory response syndrome (SIRS, n=1,578) (*24*), six
338 inflammatory/autoimmune studies (systemic lupus erythematosus (*47*), Crohn's disease, rheumatoid

339 arthritis (*48*), Ebola vaccination (*25*), neonatal-onset multisystem inflammatory disease (NOMID) and
340 macrophage activation syndrome (NLRC4-MAS) (*44*), n=326) as well as control samples from eight
341 different studies (n=408). To investigate how the COVID-19-specific co-expression modules can be linked
342 to other diseases, the combined dataset was filtered for the genes present in the COVID-19 co-expression
343 network (**Fig. 2C**) and the group fold changes were calculated across all samples (**Fig. 4B**). Additionally,
344 cell type-specific signatures (*38*) and single cell-derived neutrophil subset signatures (*42*) (**Table S6**) were
345 intersected with all CoCena[2] modules. This analysis revealed that the lightgreen module shows a high
346 (61%) neutrophil enrichment followed by module pink (38%) and maroon (32%), which is in line with a high
347 functional enrichment for neutrophil activation in lightgreen (**Fig. 2E, Table S3**). Genes within module
348 lightgreen were most prominently upregulated in the severe COVID-19 group (G1) as well as in sepsis and
349 in patients with tuberculosis and HIV infection, but not in individually occurring HIV and tuberculosis (**Fig.
350 4B**). Enrichment of the neutrophil subset signatures revealed increased expression of genes found in pre-
351 /immature neutrophils and those of inflammatory neutrophils associated with severe COVID-19. Many
352 genes within module lightgreen are known to be related to induction of neutrophil extracellular traps (NET)
353 (e.g. *PKC* (*49*), *PADI4* (*50*), *LTB4* (*51*)). Moreover, a link between excessive NET activation and tissue
354 damage has been reported in sepsis (*52*). Module darkgrey shares a similar expression pattern across the
355 disease spectrum with lightgreen and contains genes involved in platelet activation. The NET–platelet–
356 thrombin axis has been reported to be involved in the promotion of intravascular coagulation in sepsis (*53*).
357 The pink module shows the second highest neutrophil enrichment, which is dominated by the enrichment
358 of pre-/immature neutrophils subtype signatures. It is strongly increased in sepsis, tuberculosis, after Ebola
359 vaccination as well as in autoinflammatory diseases such as rheumatoid arthritis, NLRC4-MAS and NOMID,
360 and shows slight overlap with the severe COVID-19 patients in group G1. It contains many epigenetic
361 modifiers, such as *HDAC5*, *SETD1B*, or *KMT2D*, as well as *KLF2*, shown to regulate NF-κB-mediated
362 immune functions, such as inflammation, erythropoiesis and lung development (*54*). Maroon is the third
363 module with predicted neutrophil enrichment, which features genes from the "severe suppressive" subset
364 alongside the "severe inflammatory" and pre-/immature subset signatures. It is associated with COVID-19
365 groups G2-4 and shares this characteristic with blood transcriptomes from the response to infection with
366 Chikungunya and Zika virus or from HIV patients suffering from tuberculosis.
367
368 A combination of single sample gene set variation analysis (ssGSVA), a non-parametric, unsupervised
369 approach to estimate variation of gene set enrichment within each single sample, and Hallmark enrichment
370 for each disease or inflammatory condition in the compiled dataset accentuated the findings on COVID-19
371 blood transcriptomes in context of the other diseases (**Fig. 4C**). 'Interferon alpha and gamma responses'
372 were enriched in acute viral infections with Chikungunya and Zika virus as well as in HIV with or without
373 concomittent tuberculosis or after Ebola vaccination, and this enrichment was shared with COVID-19 G2.
374 'Inflammatory response', 'IL6 and TNFA signaling' is an attribute of both, G1 and G2, to a lesser degree of
375 G5, also Tuberculosis/HIV, and to some extent of sepsis. More prominently enriched in sepsis was

376 ‘complement’, ‘coagulation’, ‘heme metabolism’ and ‘glycolysis’ - shared by COVID-19 G1+G3; whereas
377 ‘oxidative phosphorylation’ and ‘mTORC1 signalling’ were seen for Chikungunya and Zika virus infections
378 - shared to some extent with COVID-19 G3+G4.
379
380 Although we observed overlaps of gene modules enriched in COVID-19 with several other infectious and
381 immune-related diseases, each of our molecularly defined COVID-19 patient groups was characterized by
382 a specific combination of these modules, clearly indicating the unique biology of this SARS-CoV-2 infection-
383 mediated immune response, which needs to be considered when developing patient-stratified therapy
384 regimens.
385

386 ## COVID-19 patient subgroup-specific signatures can be used to predict potential drug repurposing

387 Despite the immunologically-driven nature of COVID-19, most drugs that are currently investigated in
388 clinical trials to combat or ameliorate COVID-19 are targeting the virus and its direct interaction partners
389 (**Fig. 5A+S7A, Table S7**). Compounds as well as the number of clinical trials performed with anti-
390 inflammatory, immunosuppressive, and immunomodulatory properties are immensely outnumbered by
391 other approaches. Examining the listed target genes of currently investigated drugs in our stratified patient
392 groups, we found 162 included in our co-expression network analysis, most of which being differentially
393 expressed in the severe patient group G1 in comparison to G6 (**Fig. 2C+5B**). In addition, many of the
394 regulated genes in our patient signatures are clearly not affected by the drugs that are currently investigated
395 against COVID-19. The immunopathologies seen in COVID-19 patients, especially past their second week
396 of symptoms, demand a host-directed, immune system-focused therapy.

397 To identify potentially beneficial drugs, we designed an *in silico* signature-based drug repurposing approach
398 (**Fig. S7B**). To generate input signatures of interest, we characterized our stratified sample groups by
399 identifying differentially expressed genes between groups G1-G5 and the control group G6 (**Fig. S7C**).
400 Most transcriptional differences were observed for G1 (up: 4,032, down: 4,729) and G2 (up: 2,336, down:
401 2,767), whereas group G3 (up: 1,193, down: 1,921), G5 (up: 1089, down: 1216), and especially G4 (up:
402 727, down: 547) were less different to G6. Only a minor fraction of 137 DEGs was shared by all 5
403 comparisons. The most overlap of DEGs was observed between G1 and G2, the two groups comprising
404 mostly severe COVID-19 patients. Nevertheless, G2 was still characterized by a large number of specific
405 DEGs (**Fig. S7C**). GOEA of the upregulated DEGs of each comparison revealed enrichment of genes in
406 the context of ‘neutrophil activation’ and ‘coagulation’ in all comparisons (**Fig. S7D**). Humoral and B cell-
407 mediated immunity terms on the other hand were enriched the strongest in G4-specific upregulated DEGs
408 (**Fig. S7D**). Differential expression analysis for the stratified sample groups once more emphasized that
409 neutrophils play a central role in the host’s immune response against SARS-CoV-2 infection. Neutrophils,
410 as the most abundant circulating leukocytes, have become a therapeutic target of interest in multiple

411 disease settings in recent years (*55*). Two interesting target genes discussed in this context and already
412 addressed in clinical trials are CXCR2 and C5AR1. Consistent with the increased NLR in G1 and G2, we
413 observed significant upregulation of *CXCR2* and *C5AR1* in both groups (**Fig. S7E**).

414 Using patient cluster-specific DEGs as input (**Fig. S7C**, **Table S8**), we searched for compounds that evoke
415 a reverse signature in human cells via the NIH Library of Integrated Network-Based Cellular Signatures
416 (iLINCS) (*56*) and the Broad Institute's Repurposing Hub (*57*). The best counteracting signatures for each
417 comparison were combined with signatures for all currently investigated drugs and downloaded for further
418 analysis, resulting in about 63,000 signatures from 940 compounds/drugs. We performed gene set
419 enrichment analysis for all signatures against our COVID-19 patient comparisons and calculated the
420 difference of the up- and downregulated normalized enrichment score (ΔNES). A positive ΔNES indicates
421 drug signatures that reverse our COVID-19 signatures, whereas drugs with a negative ΔNES induce
422 signatures similar to the ones observed in COVID-19. Signatures were then grouped by k-means clustering
423 revealing groups of drug signatures that reverse specific patient subgroup signatures (e.g. cluster 5) or
424 those that have the highest impact on all patient subgroups (e.g. cluster 13, **Fig 5C**). Amongst the top
425 signatures in cluster 13 are methylprednisolone (ΔNES_{G1}=7.13), immunoglobulins (ΔNES_{G1}=6.62),
426 methotrexate (ΔNES_{G1}=4.21) and pevonedistat (ΔNES_{G1}=4.81) which are all under investigation
427 (clinicaltrials.gov), thereby proving that our *in silico* signature-based drug repurposing approach can indeed
428 predict drugs that have already been deemed potentially beneficial in this disease (**Fig. S7F**). Extracting
429 the leading edge of the most frequently targeted genes by the drugs included in cluster 13 revealed
430 alarmins, such as *S100A8* or *S100A6*, and *SERPINB1*, critical for neutrophil survival by protecting the cell
431 from proteases released into the cytoplasm during stress (*58–60*). Visualizing these genes in the co-
432 expression network deducted from the blood transcriptomes of our COVID-19 patient cohort identified most
433 of them as part of cluster lightgreen and maroon (**Fig. S7G**). Sample group G1-specific drug signature
434 cluster 5 also encompasses a considerable number of drugs currently being tested in clinical trials to fight
435 COVID-19 (**Fig. 5A+D, Table S9**). Interestingly, a lot of drug signatures in cluster 5 were related to female
436 hormones, such as alpha-estradiol (ΔNES_{G1}=2.83), estradiol-cypionate (ΔNES_{G1}=2.78), estriol
437 (ΔNES_{G1}=2.78), or chlormadinone acetate used in birth control pills (ΔNES_{G1}=2.74), but also for example
438 dexamethasone (ΔNES_{G1}=2.65) that was recently reported to reduce mortality in severe COVID-19 cases
439 requiring intubation, while showing no benefit for patients with milder disease courses (*61*). The most
440 frequently targeted genes within the signatures of cluster 5 included protein tyrosine kinase 2 beta (*PTK2B*),
441 playing an important role for integrin-mediated neutrophil degranulation (*62*, *63*), lysosomal protease
442 Cathepsin D (*CTSD*) expressed in neutrophils and monocytes, as well as the inflammatory mediator
443 Interleukin-1β (*IL1B*) (**Fig. 5E**). The majority of these target genes cluster in the G1-specific lightgreen and
444 pink, as well as in the maroon CoCena² modules. Drugs predicted to be effective for each module are
445 presented as a resource as supplementary information for further inspection (**Table S9**).

446 We used stratified blood transcriptomes from COVID-19 patients in an *in silico* signature-based approach
447 to identify potential drugs for therapeutic repurposing. Many of our identified hits are indeed already being
448 tested in clinical trials. Further, it became evident that, apart from common therapeutic avenues to address
449 the immune dysregulation in COVID-19 patients, there are patient groups that may benefit from treatments
450 targeting more precisely their immune phenotype and this phenotyping could be used for enrichment of
451 patient groups in clinical trials.

452

453 DISCUSSION

454 The global spread of SARS-CoV2 resulting in hundreds of thousands of COVID-19 cases urgently demands
455 a more thorough molecular understanding of the pathophysiology of the disease (*12*, *17*, *64*, *65*). While
456 vaccines are still under development (*66*–*68*), therapeutic management of the COVID-19 patients is key to
457 mitigate the clinical burden as well as to prevent deaths. It has become clear that there is great variety in
458 the occurrence of disease manifestation, and dysregulation of local and systemic immune responses have
459 been implicated in disease heterogeneity (*19*, *33*, *64*, *69*, *70*). Here, by applying classical bioinformatics
460 approaches and data-driven co-expression network analysis (CoCena²) on blood transcriptomes of COVID-
461 19 patients, we provide evidence for the existence of distinct molecular phenotypes that are not solely
462 explained by current clinical parameters. Particularly in severe COVID-19, we detected dramatic
463 transcriptional changes in the blood compartment with loss of T cell activation and concurrent gain of a
464 rather unique combination of neutrophil activation signals, which was not simply due to changes in cell
465 numbers since isolated neutrophils showed the same transcriptional changes. CoCena² allowed us to group
466 functionally related genes into 10 major transcriptional modules with distinct expression patterns across
467 five, on this basis newly defined COVID-19 patient groups, of which two (G1, G2) were related to severe
468 disease courses. While pronounced neutrophil-related alterations were observed in both subgroups of
469 severe COVID-19 patients (G1, G2), genes associated with coagulation and platelet function were mainly
470 elevated in patients with the most highly elevated number of neutrophils as measured by flow cytometry,
471 an information that was also deduced by linear support vector regression from transcriptome data.
472 Assessment of non-coding RNA species from whole blood transcriptomes also allowed the identification for
473 additional regulatory circuits. For example, we identify *CYTOR*, a lncRNA associated with granulocyte
474 survival (*36*) strongly upregulated in COVID-19 patient group G1, which was accompanied by strong
475 induction of CYTOR interactors such as *VIM* and *PIK3CB* (*37*). These findings strongly support the notion
476 that whole blood transcriptomics might not only be suitable for better understanding the systemic immune
477 response in COVID-19 patients, but can also be used to predict novel therapeutic targets involving distinct
478 pathophysiological mechanisms observed in COVID-19. In a 'reverse transcriptome approach', we used
479 the specific changes observed in the COVID-19-related transcriptional modules as the bait and searched
480 for inverse correlation in thousands of drug-based transcriptome signatures to predict potential drug

481 candidates. Most interestingly, we identified drug candidates that might be beneficial for all COVID-19
482 patients, but also candidates that might only be suitable for a subgroup of patients. Lastly, by comparing
483 the transcriptional modules identified in whole blood of COVID-19 patients, we identified unique differences
484 to other viral and bacterial infections, for which similar data were available, suggesting that blood
485 transcriptomes might also be used diagnostically or for outcome prediction in larger clinical cohorts,
486 treatment or vaccine trials in the near future.

487 Classical bioinformatic assessment of blood transcriptome data comparing defined groups, in this study
488 represented by control individuals and samples derived from either mild or severe COVID-19 patients,
489 already revealed important biology of the systemic immune response. For example, the most significantly
490 elevated transcript was *CD177*, a cell surface molecule on neutrophils, which was enhanced in both mild
491 and severe cases (**Fig. 1, S1**), CD177 has also been introduced as a hallmark for Kawasaki syndrome (*71*),
492 a syndrome that has been observed in several studies being increased in children and adolescents during
493 the SARS-CoV-2 pandemic (*72*–*74*). In acute Kawasaki syndrome, elevated expression of *CD177* was
494 associated with resistance to treatment with intravenous immunoglobulin (IVIG), a therapy in COVID-19
495 patients that is currently investigated in clinical trials around the world (9 trials, clinicaltrials.gov). Integrating
496 the assessment of CD177 into these trials might help to stratify patients and better predict individual therapy
497 outcome.

498 Hierarchical clustering of the most variable genes in the dataset already hinted towards further
499 heterogeneity among patients beyond the current clinical differentiation into mild and severe patients
500 (**Fig.1**). Indeed, co-expression network analysis in a data-driven fashion allowed us to define five patient
501 subgroups (G1-5) defined by 10 distinct transcriptional modules, which was corroborated in a second
502 independent cohort (**Fig. 2+S4**). Gene transcription observed in severe COVID-19 patients in G1 clearly
503 differed from severe G2 COVID-19 patients particularly in modules darkgrey, pink, orchid, and maroon (**Fig.
504 2C**). For example, biological mechanisms related to the darkgrey module included blood coagulation,
505 platelet activation, aggregation and degranulation, as well as cell-cell adhesion and integrin mediated
506 signaling. These are all mechanisms that are integral to several of the complications observed in a subset
507 of severe COVID-19 patients including increased disseminated intravascular coagulation (*75*), venous
508 thromboembolism (*75*, *76*), stroke (*77*), or acute cor pulmonale (*78*), further supporting the need for
509 advanced molecular subtyping of COVID-19 patients, as proposed here based on blood transcriptomes.
510 This is only one prominent example of the rich information within the new structure of molecular COVID-19
511 phenotypes that we provide here. For further inspection of the data we refer the reader to the online tool
512 that allows to extract module and group specific gene expression information
513 (https://www.fastgenomics.org/).

514 In addition to many other infectious and non-infectious diseases (*20*–*28*), whole blood transcriptomics
515 revealed important insights into the patient structure in COVID-19 and comparative analysis provides first

516 evidence for the unique changes elicited by this disease within the host in comparison to other infections
517 (**Fig. 4**). While cases in G2-4 shared changes with other viral infections such as Chikungunya or Zika,
518 mainly including interferon signature genes (*IFI16*, *IFI35*, *IFIT1*, maroon module), partial overlap to bacterial
519 sepsis was observed for G1-G3, albeit the major sepsis module (pink) was not prominently enriched in
520 COVID-19 patients indicating that there are distinct differences in pathology of these two diseases. Although
521 we could establish an integrative model using historical and publicly available blood transcriptome data, we
522 also realized that limited standardization of the experimental procedures (sample processing, library
523 production, sequencing) between different whole blood transcriptomics studies led to the exclusion of
524 several additional important studies. In this context, it will be of great interest whether blood transcriptomics,
525 as it was shown for tuberculosis (*20*, *21*), can be utilized in large enough cohorts and clinical trials for
526 disease risk or outcome prediction in COVID-19. We propose to collect whole blood transcriptomics data
527 in a central registry for direct inspection by the research community and provide a prototype model for such
528 a registry on FASTGenomics. Transcriptome data have been successfully used to predict a role for specific
529 gene networks in the drug response to certain cancer types (*79*–*83*). Considering the strong influence of
530 the systemic immune response on severity and outcome of COVID-19, we wanted to establish, whether
531 the global assessment of molecular subgroups of COVID-19 patients could be utilized to predict novel drug
532 targets for this disease addressing the dysregulated peripheral immune response of the host (**Fig. 5**). Using
533 two major databases providing transcriptome signatures to many known drugs, CLUE (*83*) and iLINCS (*82*),
534 we designed an *in silico* signature-based drug repurposing approach, allowing us to identify candidate drugs
535 (*84*) that might reverse immune pathophysiology as observed in blood transcriptomes. Some of the
536 candidate drugs identified are currently already in clinical trials, for example Imatinib (NCT04394416,
537 NCT04357613, NCT04346147, NCT04356495), Ruxolitinib (NCT04348071, NCT04355793,
538 NCT04377620) or Nintedanib (NCT04338802), for which prediction was particularly high in G1 patients.
539 These trials might benefit from assessing molecular phenotypes of immune cells thereby determining
540 whether patients with G1 type transcriptomes benefit most from such treatment. First study reports have
541 recently declared strong benefit for Dexamethasone treatment in severe COVID-19 cases requiring
542 intubation, while no effect on mortality was seen for those patients who did not require respiratory support
543 (*61*). Of note, drugs predicted to potentially reverse the transcriptome signatures of the severely affected
544 G1 group may have adverse effects in milder COVID-19 cases from G4 as observed in the contrasting
545 regulation patterns in many of the clusters (**Fig. 5C**). Interestingly and in line with the reports on sexual
546 dimorphism in COVID-19 severity and mortality (*85*), G1 included only male patients and many of the drugs
547 predicted to reverse the G1-specific signatures were related to female hormones (**Fig. 5D**). However, we
548 also predicted drugs for all COVID-19 patients already in clinical trials such as immunoglobulins (>80 trials,
549 clinicaltrials.gov), or methylprednisolone (>20 trials), findings further supporting the value of our prediction
550 approach. Despite these promising results, strongly suggesting that reverse transcriptomics is not only of
551 value in cancer (*79*–*81*) but might also be used to identify drugs targeting the immune pathophysiology in
552 COVID-19, we would also like to point out current limitations of our findings that need to be addressed in

553 future studies. Predictions will further benefit from and focused by validation studies in independent COVID-
554 19 patient cohorts, which is to be fostered by a central database for COVID-19 patients' blood transcriptome
555 data. Nevertheless, we used samples from different countries, illustrating the generalizability. Furthermore,
556 the molecularly derived and prioritized drug candidates presented here might be tested in very recently
557 introduced pre-clinical models (*86*) prior to starting clinical trials. Irrespective of the current shortcomings,
558 we favor such drug candidate identification, since it is based on interrogation of molecular data directly
559 derived from patients' immune cells involved in the ongoing processes in the disease and therefore may
560 increase the likelihood of a beneficial effect in patients.

561 Collectively, we provide first evidence for whole blood transcriptomics to potentially become a valuable tool
562 for distinguishing COVID-19 from other infections in cases for which pathogen detection might be difficult,
563 for monitoring and potentially predicting outcome of the disease, to further dissect molecular phenotypes
564 of COVID-19, particularly of the host's immune system, also along the disease course over time, and to
565 support drug target prediction for subgroups of patients. Clearly, in contrast to more sophisticated higher
566 resolution methods, whole blood transcriptomes can be easily obtained in large clinical cohort studies and
567 large clinical treatment trials yet providing an enormous information content about the molecular reactions
568 of the host's immune system. We therefore propose a blood transcriptome registry following the model we
569 introduce here on the FASTGenomics platform that would allow the scientific community to utilize the
570 information for new clinical studies and to address further large-scale studies into pathophysiological
571 mechanisms of the disease and enhance the chances of trials to demonstrate a clinical benefit in patients.

572 **Acknowledgements**
573
574 We thank Claudia Finnemann for perfect technical assistance.
575
576
577 **Deutsche COVID-19 Omics Initiative (DeCOI)**

578 Robert Bals, Alexander Bartholomäus, Anke Becker, Ezio Bonifacio, Peer Bork, Thomas Clavel, Maria
579 Colome-Tatche, Andreas Diefenbach, Alexander Dilthey, Nicole Fischer, Konrad Förstner, Julien Gagneur,
580 Alexander Goesmann, Torsten Hain, Michael Hummel, Stefan Janssen, René Kallies, Birte Kehr, Andreas
581 Keller, Sarah Kim-Hellmuth, Christoph Klein, Oliver Kohlbacher, Jan Korbel, Ingo Kurth, Markus Landthaler,
582 Yang Li, Kerstin Ludwig, Oliwia Makarewicz, Manja Marz, Alice McHardy, Christian Mertes, Markus Nöthen,
583 Peter Nürnberg, Uwe Ohler, Stephan Ossowski, Jörg Overmann, Klaus Pfeffer, Alfred Pühler, Nikolaus
584 Rajewsky, Markus Ralser, Olaf Rieß, Stephan Ripke, Ulisses Nunes da Rocha, Philip Rosenstiel, Antoine-
585 Emmanuel Saliba, Leif Erik Sander, Birgit Sawitzki, Philipp Schiffer, Joachim L. Schultze, Alexander
586 Sczyrba, Oliver Stegle, Jens Stoye, Fabian Theis, Janne Vehreschild, Jörg Vogel, Max von Kleist, Andreas
587 Walker, Jörn Walter, Dagmar Wieczorek, John Ziebuhr
588

589 **Funding**

590 ACA was supported by an intramural grant from the Department of Genomics & Immunoregulation at the
591 LIMES Institute. The work of JLS and MMBB was supported by the German Research Foundation (DFG)
592 under Germany's Excellence Strategy – EXC2151 – 390873048 as well as by the Diet-Body-Brain
593 Competence Cluster in Nutrition Research funded by the Federal Ministry of Education and Research (grant
594 numbers 01EA1410C and 01EA1809C). JLS was further supported by the DFG under SCHU 950/8-1; GRK
595 2168, TP11; SFB704, the EU project SYSCID under grant number 733100. MGN was supported by a
596 Spinoza grant of the Netherlands Organisation for Scientific Research and an ERC Advanced Grant
597 (833247). The study was funded in part by the Hellenic Institute for the Study of Sepsis. EJG-B received
598 funding from the FrameWork 7 program HemoSpec and from the Horizon2020 Marie-Curie project
599 European Sepsis Academy (granted to the National and Kapodistrian University of Athens). JN was
600 supported by the DFG (SFB TR57, SPP1937), the DZIF, and the Hector-Foundation (M89).
601

602 **Author contributions:**

603 Conceptualization, ACA, FvdV, MGN, JLS, MK, TU; Methodology, MO, MN-G, LB, NR, KB, RK, TSK, CK,
604 MH, LH, IG, SA, KD, LL, NB, MB, KH, MK, HT, SM, EDD, TU; Formal Analysis, ACA, MO, MN-G, LB, NR,
605 KB, RK, TP, TSK, CK, MH, LH, AH, IG, SA, KD, LL, NB, JG, JS-S, LS, TU; Investigation, ACA, MO, MN-
606 G, LB, NR, KB, RK, TP, TSK, CK, MH, LH, AH, IG, SA, KD, LL, NB, JS-S, JLS, TU; Resources, MM, BK,
607 NK, KG, MS, SD, NR, JR, KMK, MTV, GR, VT, AA, PP, MK, MMBB, JN, AK, EJG-B; Writing – Original

608 Draft, ACA, MO, MN-G, LB, NR, KB, RK, TP, TSK, CK, MH, LH, AH, IG, KD, MvU, AD, JLS, TU; Writing –
609 Review & Editing, ACA, BK, MO, MN-G, LB, NR, KB, RK, AD, FvdV, MGN, JLS, MK, MMBB, JN, EJG-B,
610 TU; Visualization, ACA, MO, MN-G, LB, NR, KB, RK, TP, TSK, CK, MH, LH, AH, IG, SA, KD, MvU, LL, NB,
611 JLS, TU; Supervision, ACA, JLS, MK, MMBB, JN, TU; Funding Acquisition, ACA, JLS, MMBB, EJG-B.

612

613 **Declaration of conflict of interests**

614 EJG-B has received honoraria (paid to the University of Athens) from AbbVie USA, Abbott CH, Angelini
615 Italy, Biotest Germany, InflaRx GmbH, MSD Greece and XBiotech Inc. He has received independent
616 educational grants from AbbVie, Abbott, Astellas Pharma, AxisShield, bioMérieux Inc, InflaRx GmbH and
617 XBiotech Inc.

METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Cohorts

Whole blood samples for RNA-seq analysis

The study was conducted between March 13 and March 30, 2020. A total of six ml of blood was sampled from patients with community-acquired pneumonia (CAP) by SARS-CoV-2 within the first 24 hours of hospital admission. CAP was defined as the presence of diffuse infiltrates in chest X-ray or chest computed tomography and positive molecular testing of respiratory secretions for SARS-CoV-2. Exclusion criteria were infection by the human immunodeficiency virus; neutropenia; and any previous intake of immunosuppressive medication (corticosteroids, anti-cytokine biologicals and biological response modifiers). The studies were conducted under the 23/12.08.2019 approval of the Ethics Committee of Sotiria Athens General Hospital; and the 26.02.2019 approval of the Ethics Committee of ATTIKON University General Hospital. Written informed consent was provided by patients or by first-degree relatives in case of patients unable to consent. Patients were classified as severe when they were admitted to the intensive care unit because of need of mechanical ventilation; remaining patients were hospitalized in the ward and were classified as mild. The following information was recorded: white blood cell count and differential; administered treatment; and 28-day outcome. A volume of 2.5 ml of the collected blood was transferred into one PAXgene tube and stored at -80°C. The remaining was used for flow cytometry analysis. A similar amount of blood was sampled from 10 controls fully matched for age, gender and the Charlson's comorbidity index.

For the second cohort, whole blood samples were collected for RNA-seq analysis in PAXgene tubes from 30 patients upon admission to the Intensive Care Unit of the Radboud university medical centre in Nijmegen, the Netherlands. The study was carried out in accordance with the applicable rules concerning the review of research ethics committees and informed consent. All patients or legal representatives were informed about the study details and could decline to participate. COVID-19 was diagnosed by a positive SARS-CoV-2 RT-PCR test in nasopharyngeal and throat swabs and/or by typical chest CT-scan findings. Exclusion criteria were hematological malignancies and/or active chemotherapy, solid organ transplant, auto-immune diseases, and pre-existent use of high dose corticosteroids.

Granulocyte samples for RNA-seq analysis

This study was approved by the Institutional Review board of the University Hospital Bonn (073/19 and 134/20). After providing written informed consent, 11 COVID-19 patients were included in the study. In-patients who were not able to consent at the time of study enrollment, consent was obtained after recovery.

651 COVID-19 patients who tested positive for SARS-CoV-2 RNA in nasopharyngeal swabs were recruited at
652 the Medical Clinic I of the University Hospital Bonn between March 30 and May 17, 2020.

653 Granulocytes were isolated from EDTA-treated or heparinized peripheral blood by density centrifugation
654 over Pancoll or Ficoll-Paque density centrifugation (density: 1.077g/ml). Granulocyte fractions were then
655 treated with 10ml RBC lysis buffer (Biolegend) for 10min. After RBC lysis, cells were washed with DPBS
656 and recovered by centrifugation at 300xg for 10min. Granulocyte pellets were then lysed with 500µl of
657 QIAzol (QIAGEN), shortly vortexed and incubated 5min at RT prior storage at -80°C until RNA extraction.

658

659 Rhineland Study as control samples within the integrated dataset for disease comparison

660 Study population

661 The Rhineland Study is an ongoing community-based cohort study in which all inhabitants of two
662 geographically defined areas in the city of Bonn, Germany aged 30–100 years are being invited to
663 participate. Persons living in these areas are predominantly German with Caucasian ethnicity. Participation
664 in the study is possible by invitation only. The only exclusion criterion is insufficient German language skills
665 to give informed consent.

666 Ethical Approval

667 Approval to undertake the Rhineland Study was obtained from the ethics committee of the University of
668 Bonn, Medical Faculty. The study is carried out in accordance with the recommendations of the International
669 Conference on Harmonization (ICH) Good Clinical Practice (GCP) standards (ICH-GCP). Written informed
670 consent was obtained from all participants in accordance with the Declaration of Helsinki.

671 Blood withdrawal

672 Overnight fasting blood was collected from all participants between 7:00 and 9:30 AM, including a PAXgene
673 tube for RNA extraction.

674

675 **METHOD DETAILS**

676 **Flow cytometry techniques**

677 Whole blood cells were incubated for 15 minutes in the dark with the monoclonal antibodies anti-CD14
678 FITC, anti-CD3 FITC, anti-CD4 FITC and anti-CD19 FITC (fluorescein isothiocyanate, emission 525nm,

679 Beckman Coulter); with anti-CD4 PE, anti-CD8 PE, and anti-CD(16+56) PE (phycoerythrin, emission
680 575nm, Beckman Coulter); and with anti-CD45 PC5 (emission 667nm, Beckman Coulter). Fluorospheres
681 (Beckman Coulter) were used for the determination of absolute counts. Cells were analyzed after running
682 through the CYTOMICS FC500 flow cytometer (Beckman Coulter Co, Miami, Florida). Isotypic IgG controls
683 stained also with anti-CD45 were used for each patient.

684

685 **Whole blood RNA isolation**

686 Total RNA was isolated from whole blood samples stored and stabilized in PAXgene RNA tubes using the
687 Qiagen PAXgene Blood miRNA kit according to manufacturer's guidelines. Eluted RNA was dissolved in
688 RNase free water. The quality and quantity of RNA was evaluated by visualization of 28S and 18S band
689 integrity on a Tapestation 4200 system (Agilent).

690

691 **RNA-sequencing**

692 Total RNA was converted into double-stranded cDNA libraries using the TruSeq Stranded Total RNA with
693 Ribo-Zero Globin kit (Illumina). In brief, ribosomal and globin mRNA were depleted from 750ng purified total
694 RNA using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads, remaining
695 RNA was fragmented using divalent cations under elevated temperature. First-strand was generated using
696 SuperScript2 RT (Invitrogen) supplemented with Actinomycin D, followed by second-strand synthesis with
697 dUTP replacing dTTP. 3' ends were adenylated and index adapters were ligated before subsequent PCR
698 amplification to yield the final library. Remaining overhangs were converted into blunt ends via
699 exonuclease/polymerase activities and enzymes were removed. Selective enrichment of DNA fragments
700 with ligated adaptor molecules was performed using Illumina PCR primers in a 15 cycles PCR reaction,
701 followed by purification cDNA using SPRIBeads (Beckman-Coulter). Libraries were quantified by Qubit
702 dsDNA HS Assay (Thermo Fisher Scientific) and fragment size distribution was determined using the HS
703 D1000 assay on a Tapestation 4200 system (Agilent). High-throughput sequencing was carried out with a
704 NovaSeq™ 6000 Sequencing System S2 (50bp paired-end reads), and data was converted into fastq files
705 using bcl2fastq2 v2.20.

706 **RNA-sequencing analysis**

707 Sequenced reads were aligned and quantified using STAR: ultrafast universal RNA-seq aligner (v2.7.3a)
708 (*87*) and the human reference genome, GRCh38p13, from the Genome Reference Consortium. Raw counts
709 were imported using DESeqDataSetFromHTSeqCount function from DEseq2 (v1.26.0) (*88*) and rlog
710 transformed according to DEseq2 pipeline. DESeq2 was used for the calculation of normalized counts for
711 each transcript using default parameters. All normalized transcripts with a maximum over all row mean

712 lower than 10 were excluded resulting in 37,526 present transcripts. Differentially expressed genes were
713 calculated for the scenarios status (COVID-19 vs. controls), mild/severe (severe COVID-19 vs controls,
714 mild COVID-19 vs controls, and severe vs mild COVID-19) and new_cluster (1vs6, 2vs6, 3vs6, 4vs6, and
715 5vs6) separately using a p-value cut-off of 0.05, an adjusted p-value (IHW) < 0.05 (independent hypothesis
716 weighting) and a FC of 2. All present transcripts were used as input for principal component analysis. The
717 top 25% most variable transcripts within the dataset were selected and visualized in a heat map. DEGs
718 were visualized as DE bar plots and were used as input for volcano plots.

719 Gene ontology enrichment analysis (GOEA)

720 To test for functional enrichment within all three scenarios, we performed GOEA for up- or downregulated
721 transcripts in the respective comparison using gene ontology set of biological processes. Gene set
722 "c5.bp.v7.0.symbols.gmt" was obtained from the Molecular Signatures Database (MSigDB) (*89*).
723 compareCluster and enrichGo functions from the R package ClusterProfiler (v3.12.0) (*90*) were used to
724 determine significant enrichment (q-value<0.05) of biological processes. All present genes were used as
725 background (universe).

726 Filtering for transcription factors, epigenome, surfaceome and secretome

727 All present transcripts were filtered and sorted by their variance in the dataset. The 20 most variable genes
728 of each category were selected and visualized using a heat map. Transcription factor lists were extracted
729 from (*91*), the epigenome gene list was literature-driven, surface and secretome markers were extracted
730 from the Human Protein Atlas (*92*).

731 Clustering of patients according to clinical parameters

732 The contribution of each clinical parameter to the transcriptome in COVID-19 patients was determined using
733 linear modelling of each parameter separately with PC1. Clinical parameters with rounded up adjusted r-
734 square ≥0.2 were used for agglomerative hierarchical clustering of the COVID-19 patients. A dissimilarity
735 matrix based on Gower distance was calculated using the daisy function from the cluster packages (version
736 2.1.0). Agglomerative hierarchical clustering was performed using the hclust function, defining the method
737 with a setting for ward.D2 method linkage. We evaluated the clustering by extracting clusters statistics using
738 the function cluster.stats from the package fpc (version 2.2-5). The number of clusters was chosen at the
739 value at which the lowest distance among patients within clusters (i.e. low value of within-cluster sum of
740 squares distance) and preserving a high distance among clusters (i.e. high average silhouette width) was
741 achieved, while still maintaining a comparable number of individuals among the clusters.

742 Linear support vector regression

743 Linear support vector regression (*38*) was employed to computationally deconvolute the study's whole
744 blood samples. Gene expression tables were normalized with DESeq2 and were utilized as the input
745 mixture file. LM22-subsetted signatures for B cells, T cells, NK cells, monocytes, dendritic cells, eosinophils
746 and neutrophils were generated as described on https://cibersort.stanford.edu/tutorial.php. The algorithm
747 was subsequently run with 1,000 permutations and the proportions of cell types were visualized with ggplot2
748 (v3.2.1) (*93*).

749 CoCena²: Construction of Co-expression networks, analysis - automated

750 To define differences and similarities in transcript expression patterns among the different groups, CoCena²
751 (Construction of co-expression Network Analysis – automated) was performed based on Pearson
752 correlation. CoCena² is a network-based approach to identify clusters of genes that are co-expressed in a
753 series of observed conditions based on data retrieved from RNA-sequencing. The tool offers a variety of
754 functions that allow subsequent in-depth analysis of the biological context associated with the found
755 clusters. As input for the analysis the 10,000 most variable genes were used.

756 To identify genes whose expression patterns are highly similar across all tested samples, pairwise Pearson
757 correlation coefficients are calculated using the R package Hmisc (v4.1-1). The underlying assumption of
758 the Pearson correlation to the data is that it is normally distributed, which is a valid assumption to make in
759 the context of gene expression when looking at expression patterns within different experimental conditions.
760 The correlation between each pair of genes is the basis for the subsequent network construction. Therefore,
761 the tool focuses mainly on positively correlated gene pairs, since the rate of confirmation of an edge
762 representing an association of genes is higher than that of a non-existing association.

763 In order to refine the structure of the upcoming network and to unravel the condition specific signatures, a
764 correlation cut-off is proposed to mark the minimal correlation a pair of genes must exhibit for their co-
765 expression to be taken into account. The cut-off is determined based on different criteria:

766 1) Scale-free topology:

767 Gene expression networks have been argued to have a scale-free topology (*94*), meaning that the majority
768 of vertices has a low number of adjacent edges, also referred to as the vertex' degree, whereas only very
769 few vertices have a high degree. The degree distribution of scale-free networks asymptotically follows a
770 power law. To assess the scale-free topology of a network constructed by a given correlation cut-off, a log-
771 log plot of the degree distribution is constructed and the R^2-value of the resulting linear regression is used
772 to evaluate the scale-free criterion.

773 2) Number of graph components:

774 A graph component is a subset of nodes, such that there is a path from every node within the component
775 to any other node in that same component but none connecting the nodes to any outside of that component.
776 Even though there exist functional collections of genes that cooperate to fulfil a common task, these
777 collections are not expected to be operating independently within the cell. Thus, the cut-off proposal favors
778 graphs with a small number of components.

779 3) Number of edges:

780 To avoid a highly connected graph with great lack of structure -"hairball"-, the cut-off is chosen such that
781 the number of edges is minimized while respecting the above-mentioned criteria.

782 A Pearson correlation coefficient cut-off of 0.857 (6,085 nodes and 252,584 edges) was chosen to construct
783 scale-free networks.

784 The undirected co-expression network is constructed based on the gene pairs which show a higher
785 correlation in their expression pattern than the set cut-off. A series of network-based clustering algorithms
786 is available to then identify clusters of strong co-expression within the network. An option "auto" is provided,
787 which tests the different clustering algorithms and picks the one that achieves the highest modularity score.
788 Unbiased clustering was performed using the "label propagation" algorithm in igraph (v1.2.1) [The igraph
789 software package for complex network research] and was repeated 1,000 times. Genes assigned to more
790 than 5 different clusters during the iterations received no cluster assignment.

791 To assess the expression strength of the found gene clusters in the different studied conditions, the group
792 fold changes (GFCs) of the conditions are calculated for each gene by calculating the mean expression of
793 a gene over all samples and then computing the fold change of the mean gene expression within each
794 condition from the overall mean. The GFCs of all genes within one cluster are then added and divided by
795 the total number of genes per cluster, resulting in condition-specific GFCs per cluster. Agglomerative
796 hierarchical clustering was performed by the hclust function (cluster package, version 2.1.0), using a
797 dissimilarity matrix of samples based on the GFC values of each sample defined with the daisy function for
798 calculating the Euclidean distances. The number of clusters was set to achieve a low within-cluster sum of
799 squares distance and a high average silhouette, while preserving a comparable number of individuals within
800 each cluster. The clinical parameters and the GFCs results are displayed in a heat map where conditions
801 are clustered by their GFCs revealing similar and opposing patterns (Cluster/Condition heat map).

802 Utilizing the R-package *clusterProfiler*, CoCena² automatically analyses the gene clusters with respect to
803 different kinds of gene set enrichments: The genes within each cluster are scanned for enrichment in KEGG
804 (*95*), Hallmark (*96*), Gene-Ontology terms (*97*) and Reactome (*98*). Using the R-package *pcaGoPromoter*
805 (*99*) the genes are also analyzed for enrichment of transcription factor binding sides and if the predicted

806 transcription factors are present in the data, their expression profile is visualized to facilitate evaluation of
807 their possible role.

808 To investigate the interactions between protein-coding and long-non-coding RNAs, we utilized the enricher
809 function from the clusterProfiler package. We performed an enrichment analysis for lncRNA species, using
810 the protein-coding genes that belong to the lightgreen cluster as the input gene list and all the network
811 protein-coding genes as background. The annotation table defining lncRNA to protein-coding RNA was
812 downloaded from the RNA interactome database RNAInter (*100*), filtered to only include interactions of
813 lncRNA detected by the RNA sequencing, had an experimental validation score of at least 0.5 and were
814 involved in regulating the function of granulocytes (*36*). Next, to obtain a comprehensive understanding of
815 the lncRNA that may be relevant for this specific network module, the lncRNA found by the enrichment
816 analysis with p-value <0.1 were sorted according to the highest number of genes. Thereafter, Spearman
817 correlation amongst the gene expression of each lncRNA and its corresponding protein-coding RNAs was
818 performed, and significant protein-coding RNA genes were plotted in a heat map. The $CoCena^2$ network
819 was visualized by using the ggplot function from the ggplot2 package. Annotations were generated by
820 filtering the edges of the network for the 5 top connected transcription factors, epigenetic regulators, and
821 surface and secretome markers in each cluster. GO enrichment analysis was performed on each cluster
822 by utilizing the enrichGO function from the clusterProfiler package to assess the overall functionality of the
823 cluster using the genes of each cluster as the input and all the in the network as background. The top GO
824 term and top connected genes of each cluster were compiled representing their general characteristic.

825 Granulocyte dataset analysis

826 Granulocyte raw data was aligned and quantified using STAR (v2.7.3a) and the human reference genome,
827 GRCh38p13, from the Genome Reference Consortium. Raw counts were imported using
828 DESeqDataSetFromHTSeqCount function and rlog transformed. DESeq2 was used for the calculation of
829 normalized counts for each transcript using default parameters. All normalized transcripts with a maximum
830 over all row mean lower than 10 were excluded resulting in 27,323 present transcripts. Differentially
831 expressed genes were calculated for the severe vs mild for day 1-14 and 15-28 post 1st symptoms groups)
832 separately using a p-value cut-off of 0.05, an adjusted p-value (IHW) <0.05 (independent hypothesis
833 weighting) and a FC of 2. All present transcripts were used as input for principal component analysis. DEGs
834 were visualized as DE bar plots.

835 Data Integration for Disease Comparison

836 To describe the differences and similarities between COVID-19 and other diseases, we searched in
837 databases for genomics data such as Gene Expression Omnibus (GEO) (*101*) and ArrayExpress (*102*) [for
838 studies that fulfill certain criteria: I) having at least 20 samples, II) the disease of study was of relevance
839 (other infections, such as bacterial and viral, plus diseases that mainly involve immune dysregulation, such

840 as autoimmune disease) and III) library preparation and sequencing technology differ as little as possible
841 from our COVID-19 protocol. The fastq files of 18 additional studies (588242, GSE101705, GSE107104,
842 GSE112087, GSE127792, GSE128078, GSE129882, GSE133378, GSE143507, GSE57253, GSE63042,
843 GSE66573, GSE79362, GSE84076, GSE89403, GSE90081, GSE97590, GSE99992 and the Rhineland
844 study) were downloaded and aligned with STAR. The counts were imported into R (v3.6.2) and were
845 modelled for each gene using DESeq2. Merged raw counts were filtered for the genes present in the
846 COVID-19 co-expression network, ribosomal protein-coding genes and mitochondrial genes were removed,
847 yielding a total of 5,770 genes and 3,176 samples. To account for differences in sequencing depth across
848 studies, a quantile normalization was performed on the filtered data. Group fold changes were calculated,
849 where the grouping variable was set to be the disease status.

850 To explore COVID-19 associated expression of genes within the integrated dataset, the data was
851 intersected with the gene modules previously retrieved from the COVID-19 CoCena[2] network, the mean
852 group-fold-changes were determined per cluster and condition and visualized in a heat map.
853 The modules were analyzed for enriched immune cell markers as provided by CIBERSORT and BD
854 Rhapsody and those that showed neutrophil enrichment were screened for genes representative of different
855 neutrophil subtypes as recently described (*42*).

856 Enrichment of signature from scRNA data of granulocytes

857 The signatures of different neutrophil states in COVID-19 as previously described (*42*) were enriched for
858 the different clusters from CoCena[2].

859 To get a more fine-grained differentiation of the specific neutrophil states for figure 3, the authors kindly
860 provided additional signatures from the scRNA dataset using a Wilcoxon rank sum test for differential gene
861 expression implemented in Seurat. Genes had to be expressed in >10% of the cells of a cluster, exceed a
862 logarithmic threshold >0.1 and to have >5% difference in the minimum detection between two clusters. The
863 following additional comparisons were performed: 8 and 9 (pre- and immature neutrophils combined) VS
864 the rest, 1,3,4,6 (neutrophil states from control patients) VS the rest. To get unique signature genes for
865 clusters 0, 2 and 5 (COVID-19-specific clusters) we took the following approach for each cluster: 1)
866 Calculate DEG for cluster 0 VS all other clusters. 2) Calculate DEG for cluster 0 vs 2&5. 3) Take intersection
867 of these two calculations. 4) Remove genes that occur in more than one of these intersections of cluster
868 0,2 or 5.

869 Gene set enrichment analysis (GSVA)

870 The GSVA R package (v1.34.0) (*103*) was used to test the enrichment of neutrophil signatures (*42*) in the
871 normalized gene expression table. The *gsva* method was used for the run and data were visualized in a
872 heat map with the pheatmap (v1.0.12) package.

873 ## Overview of drugs

874 An overview of currently used, recommended or investigated drugs for treatment of COVID-19 patients was
875 compiled from drug lists and lists of drugs in clinical trials downloaded from https://www.drugbank.ca/covid-
876 19, https://www.pharmgkb.org/page/COVID and https://clinicaltrials.gov/ct2/results?cond=COVID-19 (last
877 update: 2020-06-05). Classification of the drugs was performed based on the ATC code, as well as
878 additional research on the drugs action. Drug target genes were identified using the DrugBank database
879 (*104*) (**Table S6**). The number drugs currently recommended or investigated, as well as the number of
880 clinical trials within the respective drug classes were visualized using the ggplot2 package (*105*, *106*). The
881 target genes of the drugs currently recommended or investigated with a minimum frequency of 4 were
882 visualized in a word cloud using the wordcloud package (version 2.6).

883

884 ## Drug prediction

885 To identify drugs, which reverse the gene expression signature observed in the comparisons of the COVID-
886 19-specific clusters compared to the control cluster, the drug prediction databases iLINCS
887 (http://www.ilincs.org/ilincs/) and CLUE (https://clue.io/) were accessed. As input for the drug prediction the
888 top 1000 (iLINCS) or the top 100 (CLUE) DEGs were used. Drugs reversing the COVID-19 gene expression
889 signature (defined by a negative score) were pooled together with drugs under investigation in current
890 literature, resulting in a list of 940 unique drugs. Using the iLINCS API (https://github.com/uc-
891 bd2k/ilincsAPI/blob/master/usingIlincsApis.Rmd), every gene expression signature from each drug listed in
892 the signature libraries iLINCS chemical perturbagens (LINCSCP), iLINCS targeted proteomics signatures
893 (LINCSTP), Disease-related signatures (GDS), Connectivity Map signatures (CMAP), DrugMatrix
894 signatures (DM), Transcriptional signatures from EBI Expression Atlas (EBI), Cancer therapeutics response
895 signatures (CTRS) and Pharmacogenomics transcriptional signatures (PG) were downloaded. Labelling
896 was performed in the following principle: “drug name”_”database”_”database ID”. Signatures were ordered
897 by fold change and only the top 300 genes were used. This resulted in a total of 62,897 unique drug
898 signatures each with an up- and down-regulated set. Subsequently, GSEA (*107*) was performed on the
899 sequencing data for every up- and down-regulated set for each drug and each cluster comparison. The
900 resulting normalized enrichment scores (NES) were used to calculate the delta NES for each drug, defined
901 as ΔNES = NES (down) – NES (up), ergo the difference of the NES from the downregulated set and the
902 NES from the upregulated set of each respective drug. These ΔNES values were then k-mean clustered
903 (k=40). The clusters showing the highest ΔNES values for all comparisons and the cluster showing only
904 high ΔNES in the comparison G1vsG6 (most severe) were chosen and selected ones of the uniquely
905 present drugs shown. The leading edge genes of the downregulation signatures of these drugs for the

906 G1vsG6 comparison were examined and the frequency was counted. Recurring target genes were plotted
907 on the CoCena[2] network.

908 Patterns of differential gene expression of genes targeted by drugs which are currently approved or under
909 investigation for the treatment of COVID-19 patients were visualized using ggplot2. To this end, target
910 genes of each drug and their first degree neighbors were extracted from several databases and the gene
911 co-expression networks, respectively. Regulation patterns of expression of these genes in different COVID-
912 19 patient groups, as compared to the control group, were classified as up-/downregulated or not significant
913 (n.s.) when pairwise comparisons of gene expression of COVID-19 patients and controls were not
914 statistically significant. The same methodology was applied to genes not included in the drug-target list to
915 identify genes which are not targeted by current drugs but could be potentially targeted by newly identified
916 drugs.

917

918 CONTACT FOR REAGENT AND RESOURCE SHARING

919 For further information and requests for resources and reagents should be directed to and will be fulfilled
920 by the lead contact, Dr. Thomas Ulas (t.ulas@uni-bonn.de).

921

922 DATA AND SOFTWARE AVAILABILITY

923 Data Availability

924 The data that support the findings of this study, including transcriptome data from 60 patients at multiple
925 time points who granted informed consent to share such data, are made available at the European
926 Genome-Phenome Archive (EGA) under accession number EGAS00001004503, which is hosted by the
927 EBI and the CRG. The Rhineland Study's dataset is not publicly available because of data protection
928 regulations. Access to data can be provided to scientists in accordance with the Rhineland Study's Data
929 Use and Access Policy. Requests for further information or to access the Rhineland Study's dataset should
930 be directed to RS-DUAC@dzne.de. In addition to data deposition on EGA, we provide an interactive
931 platform for data inspection and analysis via FASTGenomics (fastgenomics.org). The FASTGenomics
932 platform also provides normalized count tables of the datasets generated in this study. Materials, code, and
933 data are available from the corresponding author upon reasonable request. CoCena[2] is also available under
934 https://github.com/Ulas-lab/CoCena2.

935 # REFERENCES

936 1. N. D. Grubaugh, M. E. Petrone, E. C. Holmes, We shouldn't worry when a virus mutates during
937 disease outbreaks. *Nat. Microbiol.* **5** (2020), pp. 529–530.

938 2. J. R. Fauver *et al.*, *Cell*. **181** (2020), doi:10.1016/j.cell.2020.04.021.

939 3. P. Zhou *et al.*, *Nature*. **579**, 270–273 (2020).

940 4. C. Brignola *et al.*, *J. Clin. Gastroenterol.* **10**, 631–634 (1988).

941 5. W. Guan *et al.*, *N. Engl. J. Med.* **382**, 1708–1720 (2020).

942 6. C. Huang *et al.*, *Lancet*. **395**, 497–506 (2020).

943 7. F. Zhou *et al.*, *Lancet*. **395**, 1054–1062 (2020).

944 8. D. Wang *et al.*, *JAMA - J. Am. Med. Assoc.* **323**, 1061–1069 (2020).

945 9. E. Z. Ong *et al.*, *Cell Host Microbe*. **27** (2020), doi:10.1016/j.chom.2020.03.021.

946 10. B. Wang, R. Li, Z. Lu, Y. Huang, *Aging (Albany. NY).* **12**, 6049–6057 (2020).

947 11. W. J. Guan *et al.*, *Eur. Respir. J.* **55** (2020), doi:10.1183/13993003.00547-2020.

948 12. R. T. Gandhi, J. B. Lynch, C. del Rio, *N. Engl. J. Med.* (2020), doi:10.1056/nejmcp2009249.

949 13. S. A. Dugger, A. Platt, D. B. Goldstein, Drug development in the era of precision medicine. *Nat.*
950 *Rev. Drug Discov.* **17** (2018), pp. 183–196.

951 14. A. Zumla *et al.*, Towards host-directed therapies for tuberculosis. *Nat. Rev. Drug Discov.* **14**
952 (2015), pp. 511–512.

953 15. A. Mullard, Coordinating the COVID-19 pipeline. *Nat. Rev. Drug Discov.* **19** (2020), p. 298.

954 16. H. Ledford, *Nature*. **581** (2020), doi:10.1038/d41586-020-01367-9.

955 17. D. A. Berlin, R. M. Gulick, F. J. Martinez, *N. Engl. J. Med.* (2020), doi:10.1056/nejmcp2009575.

956 18. G. Dimopoulos *et al.*, *Cell Host Microbe* (2020), doi:10.1016/j.chom.2020.05.007.

957 19. Y. Jamilloux *et al.*, Should we stimulate or suppress immune responses in COVID-19? Cytokine
958 and anti-cytokine interventions. *Autoimmun. Rev.* **19** (2020), , doi:10.1016/j.autrev.2020.102567.

959 20. D. E. Zak *et al.*, *Lancet*. **387**, 2312–2322 (2016).

960 21. E. G. Thompson *et al.*, *Tuberculosis*. **107**, 48–58 (2017).

961 22. S. Leong *et al.*, *Tuberculosis*. **109**, 41–51 (2018).

962 23. S. Verma *et al.*, *BMC Infect. Dis.* **18** (2018), doi:10.1186/s12879-018-3127-4.

963 24. E. L. Tsalik *et al.*, *Genome Med.* **6** (2014), doi:10.1186/s13073-014-0111-5.

964 25. A. Rechtien *et al.*, *Cell Rep*. **20**, 2251–2261 (2017).

965 26. D. Michlmayr *et al.*, *Mol. Syst. Biol.* **14** (2018), doi:10.15252/msb.20177862.

966 27. J. A. Hill *et al.*, *J. Virol.* **93** (2018), doi:10.1128/jvi.01419-18.

967 28. E. Bartholomeus *et al.*, *J. Transl. Med.* **17** (2019), doi:10.1186/s12967-019-2037-6.

968 29. X. Yang *et al.*, *Lancet Respir. Med.* **8**, 475–481 (2020).

969 30. G. Chen *et al.*, *J. Clin. Invest.* **130**, 2620–2629 (2020).

970 31. E. J. Giamarellos-Bourboulis *et al.*, *Cell Host Microbe* (2020), doi:10.1016/j.chom.2020.04.009.

971 32. J. Hadjadj *et al.*, *medRxiv*, in press, doi:10.1101/2020.04.19.20068015.

972 33. M. Merad, J. C. Martin, Pathological inflammation in patients with COVID-19: a key role for
973 monocytes and macrophages. *Nat. Rev. Immunol.* **20** (2020), , doi:10.1038/s41577-020-0331-4.

974 34. P. Lalezari, G. B. Murphy, F. H. Allen, *J. Clin. Invest.* **50**, 1108–1115 (1971).

975 35. R. Grieshaber-Bouyer, P. A. Nigrovic, Neutrophil heterogeneity as therapeutic opportunity in
976 immune-mediated disease. *Front. Immunol.* **10** (2019), , doi:10.3389/fimmu.2019.00346.

977 36. X. Tian, J. Tian, X. Tang, J. Ma, S. Wang, *J. Hematol. Oncol.* **9** (2016), doi:10.1186/s13045-016-
978 0333-7.

979 37. Y. Zhang *et al.*, *Biomed. Pharmacother.* **94**, 644–651 (2017).

980 38. A. M. Newman *et al.*, *Nat. Methods*. **12**, 453–457 (2015).

981 39. W. Shang *et al.*, *J. Med. Virol.* (2020), doi:10.1002/jmv.26031.

982 40. X. Yan *et al.*, *J. Med. Virol.* (2020), doi:10.1002/jmv.26061.

983 41. C. Wenham, J. Smith, R. Morgan, COVID-19: the gendered impacts of the outbreak. *Lancet*. **395**
984 (2020), pp. 846–848.

985 42. J. Schulte-Schrepping *et al.*, *medRxiv*, in press, doi:10.1101/2020.06.03.20119818.

986 43. N. Romberg *et al.*, *Nat. Genet.* **46**, 1135–1139 (2014).

987 44. S. W. Canna *et al.*, *Nat. Genet.* **46**, 1140–1146 (2014).

988 45. D. Michlmayr *et al.*, *Cell Rep.* **31** (2020), doi:10.1016/j.celrep.2020.107569.

989 46. L. S. de Araujo *et al.*, *Front. Microbiol.* **7** (2016), doi:10.3389/fmicb.2016.01586.

990 47. W. A. Figgett *et al.*, *Clin. Transl. Immunol.* **8** (2019), doi:10.1002/cti2.1093.

991 48. K. Shchetynsky *et al.*, *Arthritis Res. Ther.* **19** (2017), doi:10.1186/s13075-017-1220-5.

992 49. R. D. Gray *et al.*, *J. Inflamm. (United Kingdom)*. **10** (2013), doi:10.1186/1476-9255-10-12.

993 50. A. S. Rohrbach, D. J. Slade, P. R. Thompson, K. A. Mowen, Activation of PAD4 in NET formation.
994 *Front. Immunol.* **3** (2012), , doi:10.3389/fimmu.2012.00360.

995 51. A. Carestia *et al.*, *J. Leukoc. Biol.* **99**, 153–162 (2016).

996 52. N. L. Denning, M. Aziz, S. D. Gurien, P. Wang, Damps and nets in sepsis. *Front. Immunol.* **10**
997 (2019), , doi:10.3389/fimmu.2019.02536.

998 53. B. McDonald *et al.*, *Blood*. **129**, 1357–1367 (2017).

999 54. P. Jha, H. Das, *Int. J. Mol. Sci.* **18** (2017), doi:10.3390/ijms18112383.

1000 55. T. Németh, M. Sperandio, A. Mócsai, Neutrophils as emerging therapeutic targets. *Nat. Rev. Drug*
1001 *Discov.* **19** (2020), pp. 253–275.

1002 56. M. Pilarczyk *et al.*, *bioRxiv*, 826271 (2019).

1003 57. S. M. Corsello *et al.*, The Drug Repurposing Hub: A next-generation drug library and information
1004 resource. *Nat. Med.* **23** (2017), pp. 405–408.

1005 58. M. Baumann, C. T. N. Pham, C. Benarafa, *Blood*. **121**, 3900–3907 (2013).

1006 59. A. Torriglia, E. Martin, I. Jaadane, The hidden side of SERPINB1/Leukocyte Elastase Inhibitor.
1007 *Semin. Cell Dev. Biol.* **62** (2017), pp. 178–186.

1008 60. L. Duplomb *et al.*, *J. Mol. Med.* **97**, 633–645 (2019).

1009 61. "Low-cost dexamethasone reduces death by up to one third in hospitalised patients with severe
1010 respiratory complications of COVID-19" (2020), (available at
1011 https://www.recoverytrial.net/files/recovery_dexamethasone_statement_160620_v2final.pdf).

1012 62. M. Fuortes, M. Melchior, H. Han, G. J. Lyon, C. Nathan, *J. Clin. Invest.* **104**, 327–335 (1999).

1013 63. L. A. Kamen, J. Schlessinger, C. A. Lowell, *J. Immunol.* **186**, 1656–1665 (2011).

1014 64. W. C. Koff, M. A. Williams, *N. Engl. J. Med.* (2020), doi:10.1056/nejmp2006761.

1015 65. L. Bao *et al.*, *Nature* (2020), doi:10.1038/s41586-020-2312-y.

1016 66. N. Lee, A. McGeer, The starting line for COVID-19 vaccine development. *Lancet*. **395** (2020), pp.
1017 1815–1816.

1018 67. N. Lurie, M. Saville, R. Hatchett, J. Halton, Developing covid-19 vaccines at pandemic speed. *N.*
1019 *Engl. J. Med.* **382** (2020), pp. 1969–1973.

1020 68. E. Callaway, *Nature*. **580**, 576–577 (2020).

1021 69. K. Subbarao, S. Mahanty, *Immunity*. **52**, 905–909 (2020).

1022 70. N. Vabret *et al.*, *Immunity* (2020), doi:10.1016/j.immuni.2020.05.002.

1023 71. Y. H. Huang, M. H. Lo, X. Y. Cai, S. F. Liu, H. C. Kuo, *Pediatr. Rheumatol.* **17** (2019),
1024 doi:10.1186/s12969-019-0315-8.

1025 72. J. Toubiana *et al.*, *BMJ*. **369**, m2094 (2020).

1026 73. R. M. Viner, E. Whittaker, Kawasaki-like disease: emerging complication during the COVID-19
1027 pandemic. *Lancet*. **395** (2020), , doi:10.1016/S0140-6736(20)31129-6.

1028 74. G. Ronconi *et al.*, *J. Biol. Regul. Homeost. Agents*. **34** (2020), doi:10.23812/EDITORIAL-
1029 RONCONI-E-59.

1030 75. H. Al-Samkari *et al.*, *Blood* (2020), doi:10.1182/blood.2020006520.

1031 76. F. A. Klok *et al.*, *Thromb. Res.* **191** (2020), doi:10.1016/j.thromres.2020.04.013.

1032 77. T. J. Oxley *et al.*, *N. Engl. J. Med.* **382**, e60 (2020).

1033 78. C. Creel-Bulos *et al.*, Acute cor pulmonale in critically ill patients with covid-19. *N. Engl. J. Med.*
1034 **382** (2020), , doi:10.1056/NEJMc2010459.

1035 79. J. W. Tyner *et al.*, *Nature*. **562**, 526–531 (2018).

1036 80. F. Rambow *et al.*, *Cell*. **174**, 843-855.e19 (2018).

1037 81. Z. Qiu *et al.*, *Cancer Cell*. **36**, 179-193.e11 (2019).

1038 82. A. B. Keenan *et al.*, The Library of Integrated Network-Based Cellular Signatures NIH Program:
1039 System-Level Cataloging of Human Cells Response to Perturbations. *Cell Syst.* **6** (2018), pp. 13–
1040 24.

1041 83. A. Subramanian *et al.*, *Cell*. **171**, 1437-1452.e17 (2017).

1042 84. Y. Zhou *et al.*, *Cell Discov.* **6** (2020), doi:10.1038/s41421-020-0153-3.

1043 85. E. P. Scully, J. Haverfield, R. L. Ursin, C. Tannenbaum, S. L. Klein, Considering how biological
1044 sex impacts immune responses and COVID-19 outcomes. *Nat. Rev. Immunol.* (2020), ,
1045 doi:10.1038/s41577-020-0348-8.

1046 86. J. Shi *et al.*, *Science*. **368**, 1016–1020 (2020).

1047 87. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. **29** (2013), pp. 15–21.

1048 88. M. I. Love, W. Huber, S. Anders, *Genome Biol.* **15** (2014), doi:10.1186/s13059-014-0550-8.

1049 89. A. Liberzon *et al.*, *Bioinformatics*. **27**, 1739–40 (2011).

1050 90. G. Yu, L. G. Wang, Y. Han, Q. Y. He, *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

1051 91. D. L. Fulton *et al.*, *Genome Biol.* **10** (2009), doi:10.1186/gb-2009-10-3-r29.

1052 92. M. Uhlén *et al.*, *Science (80-.).* **347** (2015), doi:10.1126/science.1260419.

1053 93. K. Ito, D. Murphy, *CPT Pharmacometrics Syst. Pharmacol.* **2** (2013), doi:10.1038/psp.2013.56.

1054 94. P. Aloy, R. B. Russell, *EMBO Rep.* **5**, 349–350 (2004).

1055 95. M. Kanehisa, S. Goto, *Nucleic Acids Res.* **28**, 27–30 (2000).

1056 96. A. Liberzon *et al.*, *Cell Syst.* **1**, 417–425 (2015).

1057 97. Gene Ontology Consortium, *Nucleic Acids Res.* **43**, D1049-56 (2015).

1058 98. A. Fabregat *et al.*, *Nucleic Acids Res.* **46**, D649–D655 (2018).

1059 99. M. Hansen *et al.*, *PLoS One*. **7** (2012), doi:10.1371/journal.pone.0032394.

1060 100. Y. Lin *et al.*, *Nucleic Acids Res.* **48**, D189–D197 (2020).

1061 101. E. Clough, T. Barrett, in *Methods in Molecular Biology* (Humana Press Inc., 2016), vol. 1418, pp.
1062 93–110.

1063 102. A. Athar *et al.*, *Nucleic Acids Res.* **47**, D711–D715 (2019).

1064 103. S. Hänzelmann, R. Castelo, J. Guinney, *BMC Bioinformatics*. **14** (2013), doi:10.1186/1471-2105-
1065 14-7.

1066 104. D. S. Wishart *et al.*, *Nucleic Acids Res.* **46**, D1074–D1082 (2018).

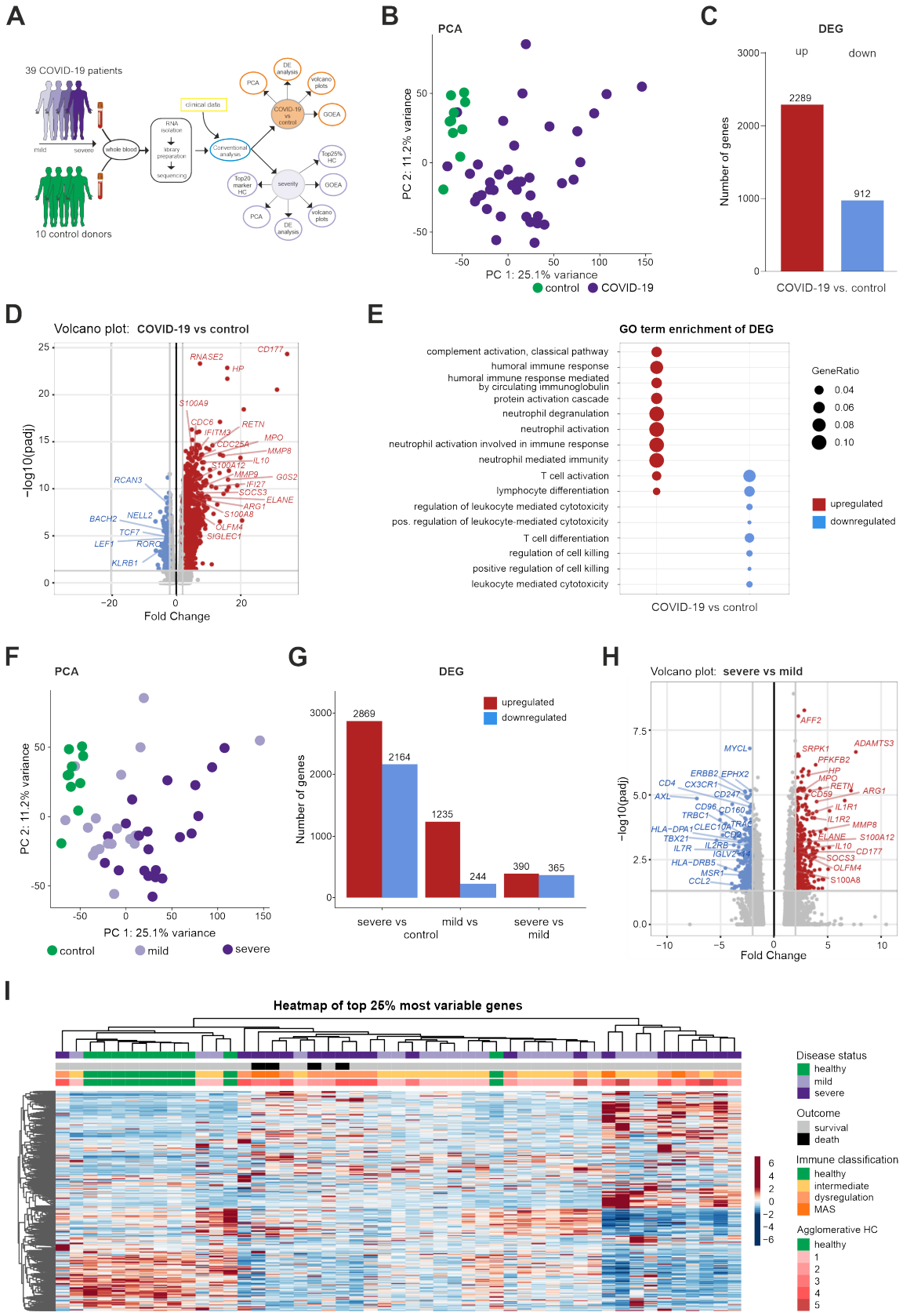
1067 105. H. Wickham, *Ggplot2 : elegant graphics for data analysis* (Springer, 2009).

1068 106. N. Neth, H.; Gradwohl, unikn: Graphical elements of the University of Konstanz's corporate
1069 design. *Soc. Psychol. Decis. Sci.* (2019), (available at https://www.spds.uni-
1070 konstanz.de/publication-page/unikn-graphical-elements-university-konstanzs-corporate-design).

1071 107. G. Korotkevich, V. Sukhov, A. Sergushichev, *bioRxiv*, 060012 (2019).

FIGURES

1073 # Figure 1



1074 *Figure 1 – Whole blood transcriptomes reveal diversity of COVID-19 patients not explained by*
1075 *disease severity*

1076 **(A)** Schematic workflow for analysis of whole blood transcriptome data.

1077 **(B)** PCA plot depicting relationship of all samples based on dynamic gene expression of all genes
1078 comparing COVID-19 and control samples.

1079 **(C)** Number of significantly upregulated (red) and downregulated (blue) genes (FC>|2|, FDR-adj. p-value
1080 <0.05) comparing COVID-19 and control samples.

1081 **(D)** Volcano plot depicting fold changes (FC) and FDR-adjusted p-values comparing COVID-19 and control
1082 samples. Differentially expressed up- (red) and downregulated genes (blue) are shown and selected genes
1083 are highlighted.

1084 **(E)** Plot of top 10 most enriched GO terms for significantly up- and downregulated genes, showing ratio of
1085 significantly regulated genes within enriched GO terms (GeneRatio).

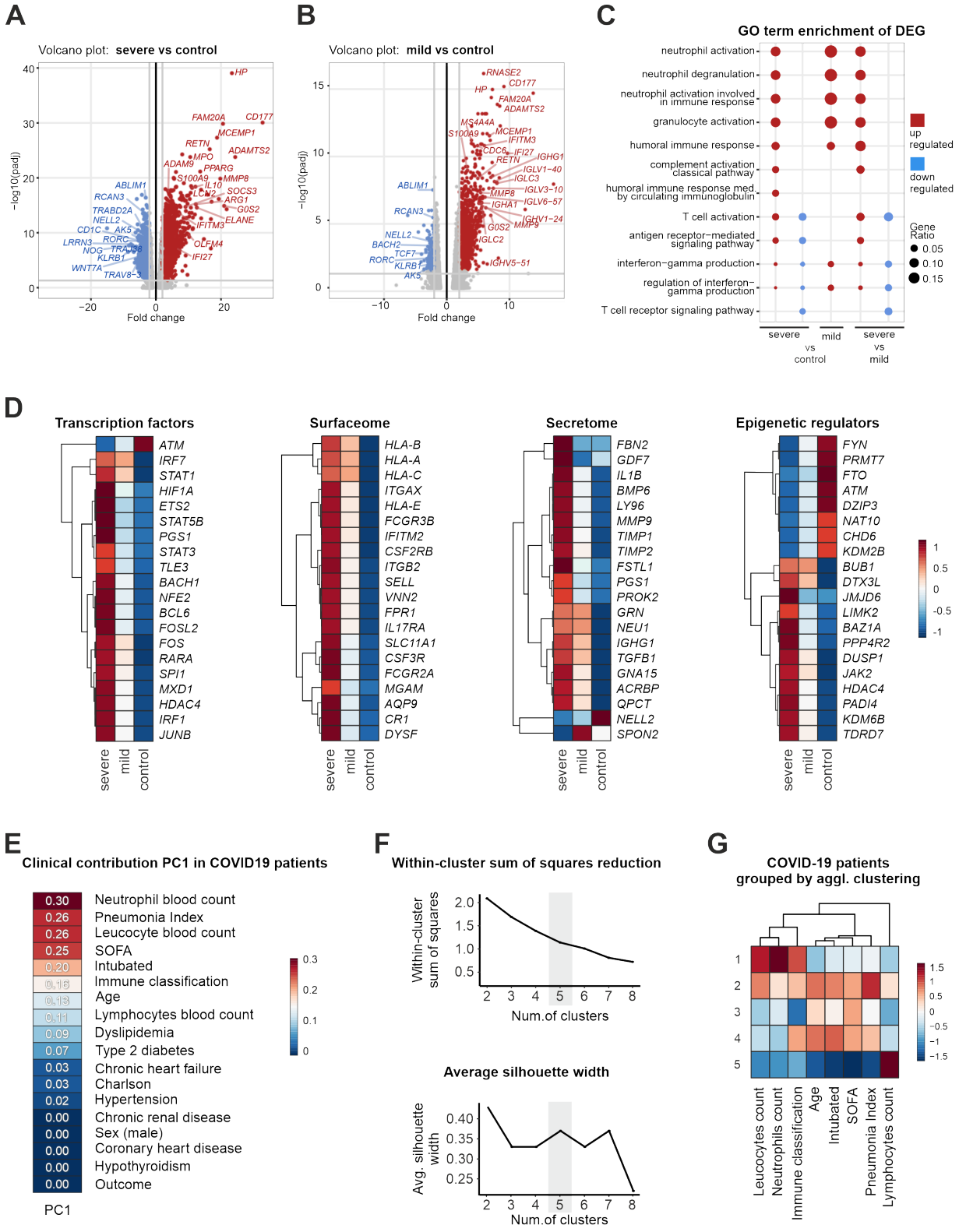
1086 **(F)** PCA plot depicting relationship of all samples based on dynamic gene expression of all genes
1087 comparing mild and severe COVID-19 as well as control samples.

1088 **(G)** Number of significantly upregulated (red) and downregulated (blue) genes (FC >|2|, FDR-adj. p-value
1089 < 0.05) comparing mild and severe COVID-19 as well as control samples.

1090 **(H)** Volcano plot depicting fold changes and FDR-adjusted p-values comparing mild and severe COVID-19
1091 as well as control samples. Differentially expressed up- (red) and downregulated genes (blue) are shown
1092 and selected genes are highlighted.

1093 **(I)** Hierarchical clustering map of 25% most variable genes between control patients, COVID-19 mild or
1094 severe patients, with additional annotation of disease outcome, hierarchical agglomerative clustering of
1095 clinical parameters COVID-19, and the groups defined by agglomerative clustering.

Supplemental Figure 1 - related to Figure 1



1096

1097 *Figure S1 – related to Figure 1*

1098 **(A-B)** Volcano plots depicting fold changes and FDR-adjusted p-values comparing severe **(A)** or mild **(B)**
1099 COVID-19 patients vs. controls. Differentially expressed up- (red) and downregulated genes (blue) are
1100 shown and selected genes are highlighted.

1101 **(C)** Plot of top 10 most enriched GO terms for significantly up- and downregulated genes. Ratios of
1102 significantly regulated genes within enriched GO terms (GeneRatio) are shown for the comparisons
1103 between mild or severe COVID-19 patients and controls as well as between 'mild' and 'severe' COVID-19
1104 patients.

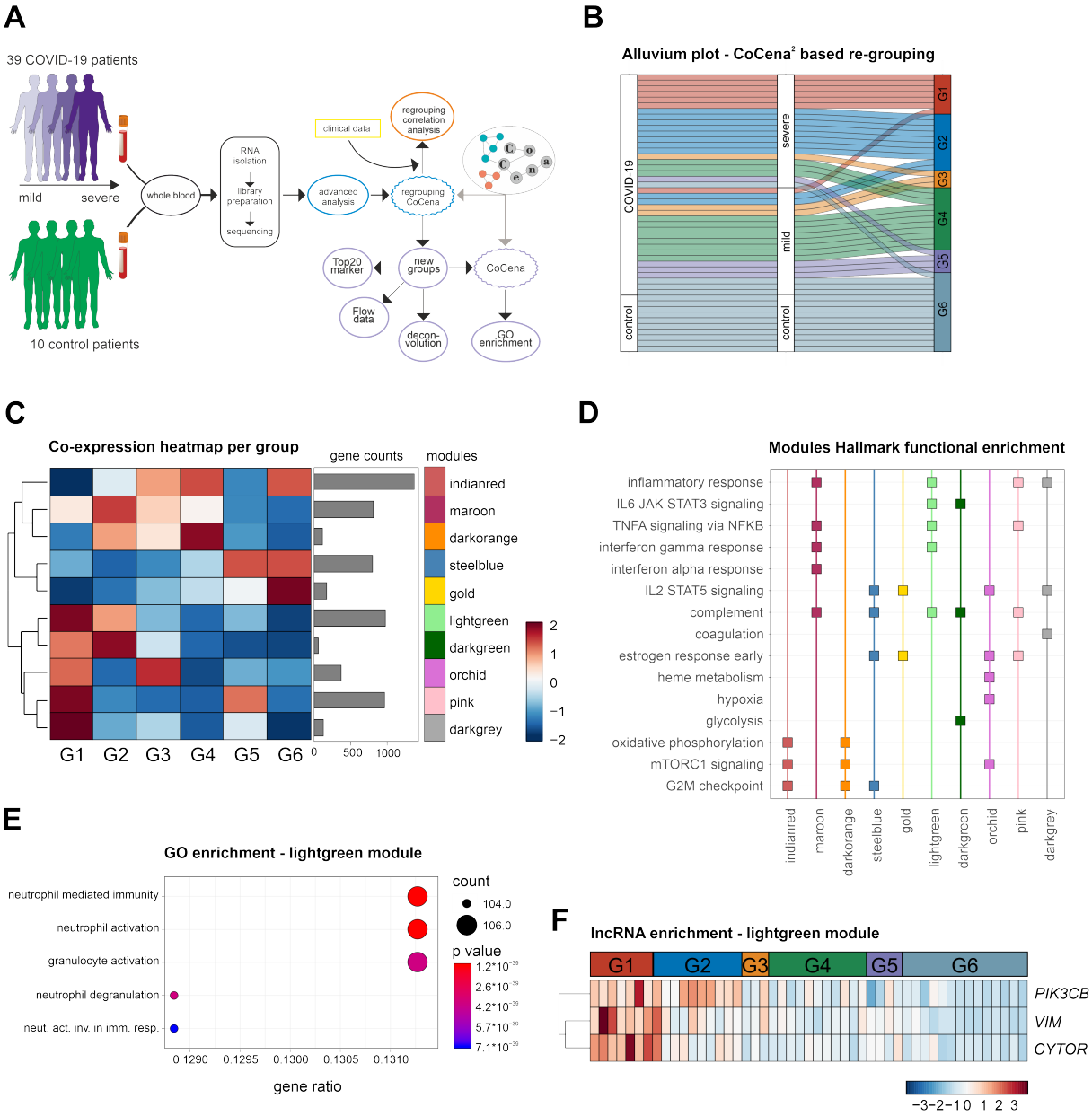
1105 **(D)** Heat map of group mean gene expression values from the top 20 most variant transcription factors,
1106 epigenetic regulators, surface and secreted proteins.

1107 **(E)** Heat map of the linear model adjusted r-square that includes each clinical parameter with PC1. Clinical
1108 parameters with r-adjusted square ≥0.1 were used for agglomerative clustering of COVID-19 patients.

1109 **(F)** Plots for agglomerative clustering statistics: within cluster sum of squares and high average silhouette
1110 width scores.

1111 **(G)** Heat map presenting summary statistics of the clinical parameters used for the clustering across clinical
1112 agglomerative clusters 1-5.

Figure 2



1113

1114 *Figure 2 – Co-expression analysis discloses COVID-19 subgroups with distinct molecular*
1115 *signatures*

1116 **(A)** Schematic overview of the analysis performed on the whole blood samples.

1117 **(B)** Alluvium plot visualizing the distribution of the samples according to different grouping; disease status,
1118 severity and data-driven sample groups.

1119 **(C)** Group fold change heat map and hierarchical clustering for the six data-driven sample groups and the
1120 gene modules identified byCoCena^2 analysis.

1121 **(D)** Functional enrichment of CoCena^2-derived modules using the Hallmark gene set database. Selected
1122 top terms were visualized.

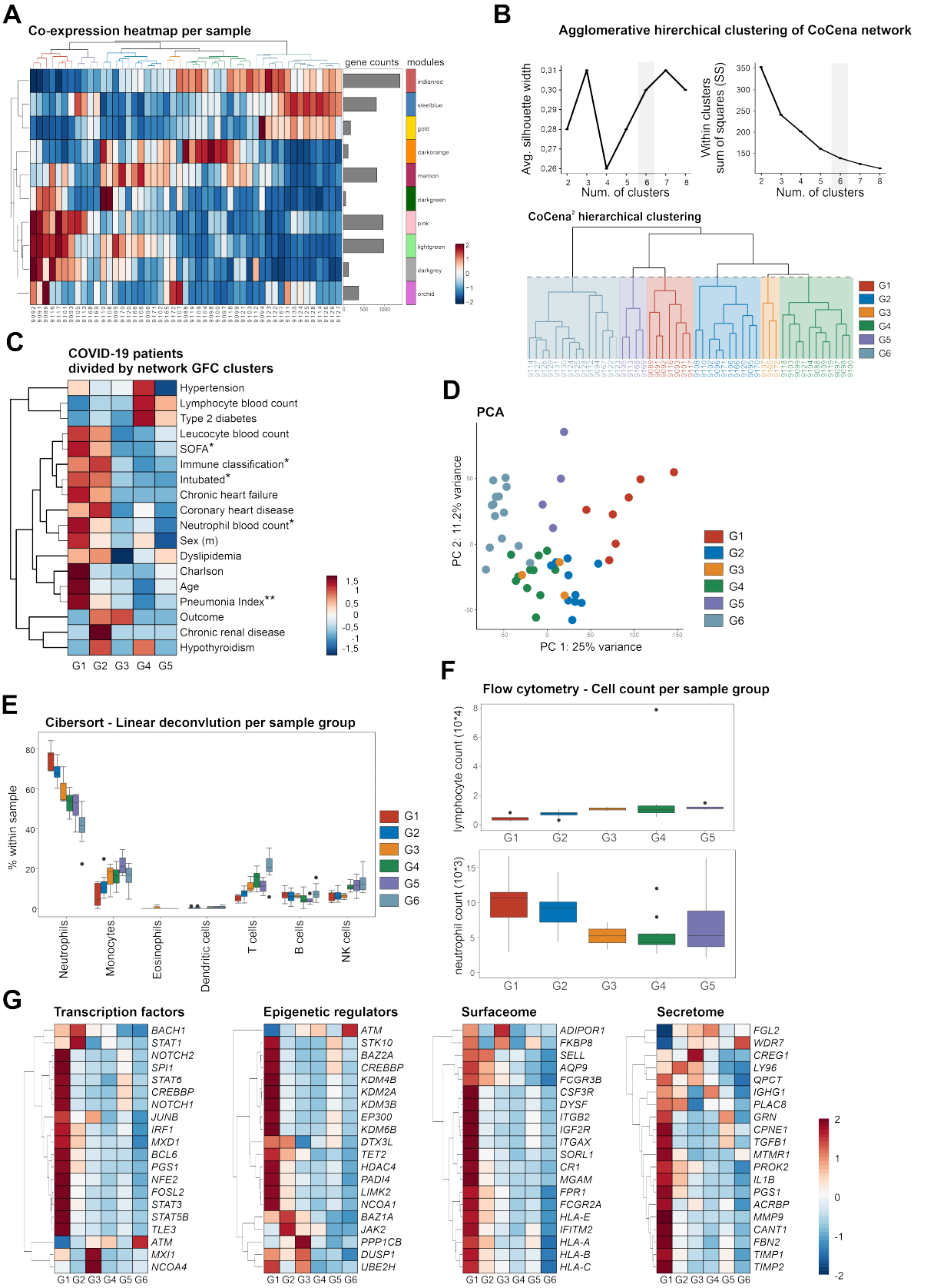
1123 **(E)** Functional enrichment of CoCena^2 module lightgreen using GO gene set database. Top 5 terms were
1124 visualized.

1125 **(F)** Heat map presenting the normalized expression values of the lncRNA *CYTOR*, and protein coding
1126 RNAs *PIK3CB* and *VIM* from the lightgreen CoCena^2 module.

1127 **(G)** Neutrophil-lymphocyte ratio plot after cell type deconvolution at lineage level.

1128 **(H)** Neutrophil-lymphocyte ratio across the six data-driven sample groups. Box plots show median with
1129 variance, with lower and upper hinges representing the 25th and 75th percentile, respectively.

Supplemental Figure 2 - related to Figure 2



1131 *Figure S2 – related to Figure 2*

1132 **(A)** Group fold change (GFC) heat map and hierarchical clustering for each sample and the gene modules
1133 identified by CoCena[2] analysis.

1134 **(B)** Agglomerative hierarchical clustering of the samples according to the GFC. Top plots present the
1135 clustering statistics (within cluster sum of squares and high average silhouette width scores) used for the
1136 generation of the six data-driven CoCena[2] sample groups G1-G6, which are plotted in the dendrogram plot.

1137 **(C)** Heat map presenting summary statistics of clinical parameters for COVID-19 patients grouped
1138 according to the CoCena[2] sample groups G1-G5 Presented are scaled values of the mean value of the
1139 parameters age/blood cell counts/SOFA score/Pneumonia Index/Charlson score, and prevalence of the
1140 comorbidity/death (outcome)/male (sex)/immune classification (intermediate/dysregulation/MAS).
1141 Statistical differences were estimated among the groups via the one-sided Anova test or Fisher test, for
1142 numeric or categorical values respectively. (*,** p-value < 0.05, 0.01 respectively)

1143 **(D)** PCA plot depicting relationship of all samples based on dynamic gene expression of all genes. Coloring
1144 based on the six data-driven CoCena[2] sample groups G1-G6.

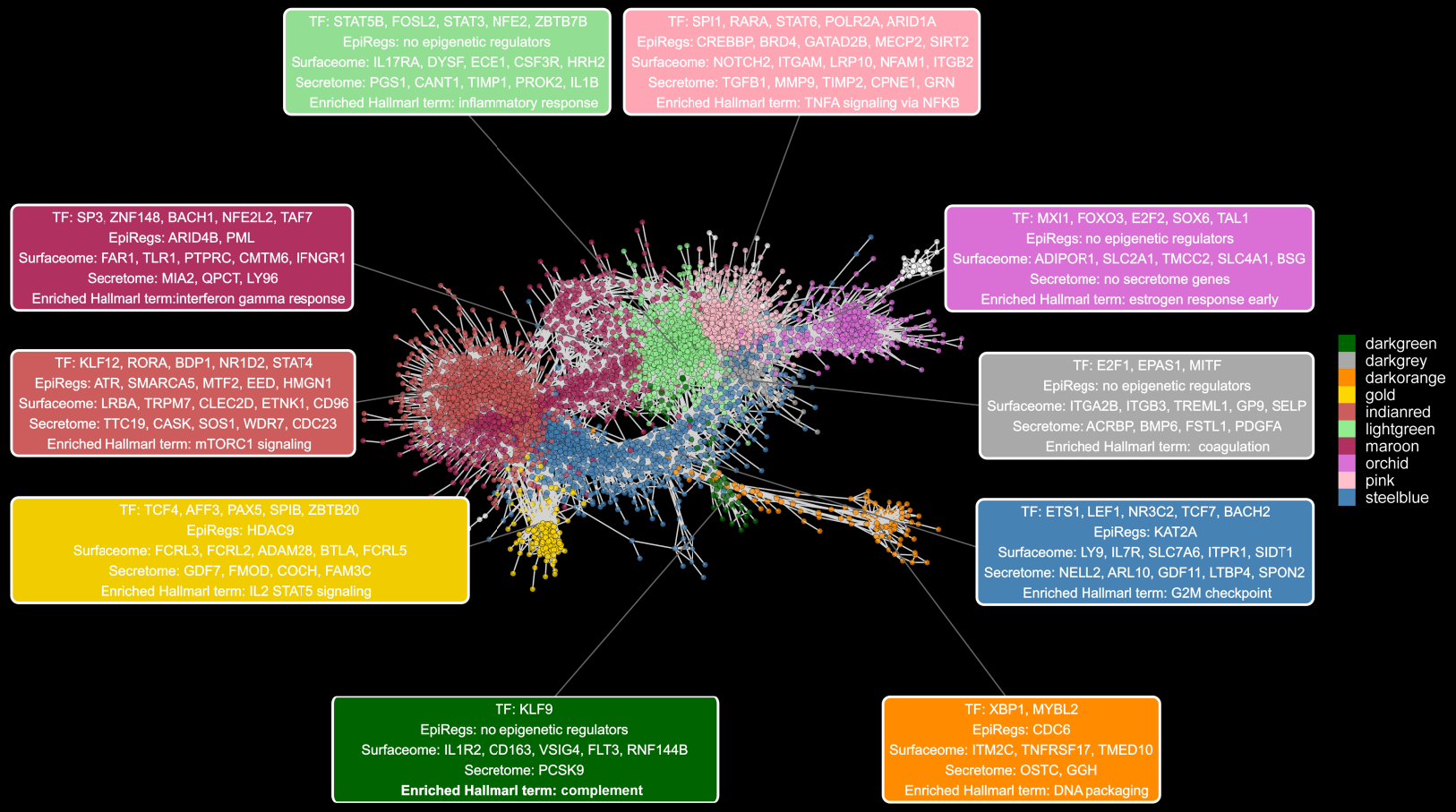
1145 **(E)** Cibersort cell type deconvolution at cell subset level. Grouping based on the six data-driven CoCena[2]
1146 sample groups G1-G6.

1147 **(F)** Flow cytometry analysis, number of lymphocytes (upper) and neutrophils (lower) per µl of blood.
1148 Grouping based on the six data-driven CoCena[2] sample groups G1-G6.

1149 **(G)** Heat map of DE and top 20 most variable transcription factors, epigenetic regulators, surface and &
1150 secreted proteins.

Supplemental Figure 3 - related to Figure 2

A



1152 *Figure S3 – related to Figure 2*

1153 **(A)** Visualization of the COVID-19 CoCena[2] network. Nodes are genes and edges represent co-expressed
1154 genes. Additional module information is displayed by module-colored labels. Labels include information
1155 about top-connected transcription factors (TFs), epigenetic regulators, surface & secreted markers as well
1156 as representative Hallmark terms.

Supplemental Figure 4 - related to Figure 2

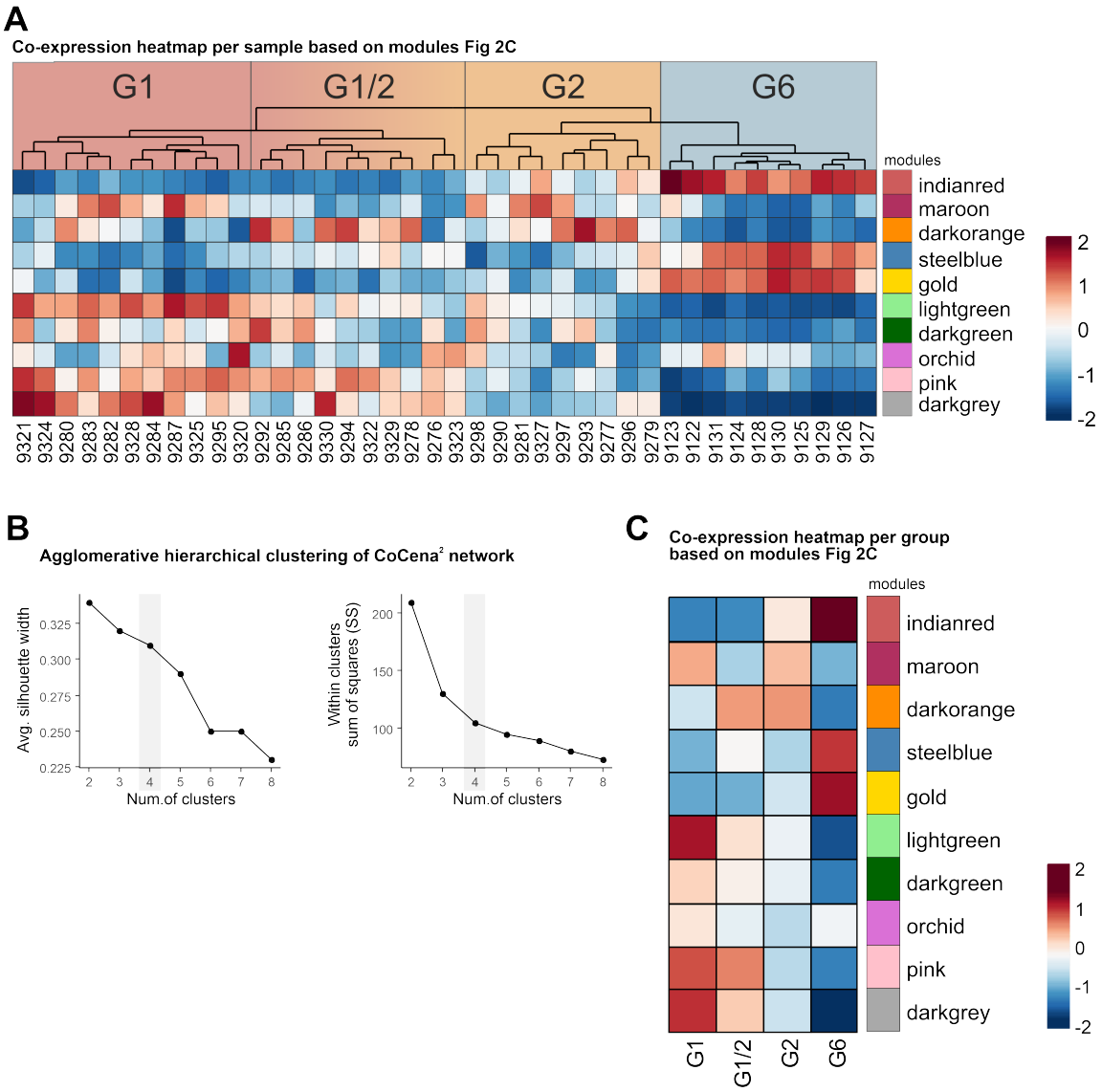


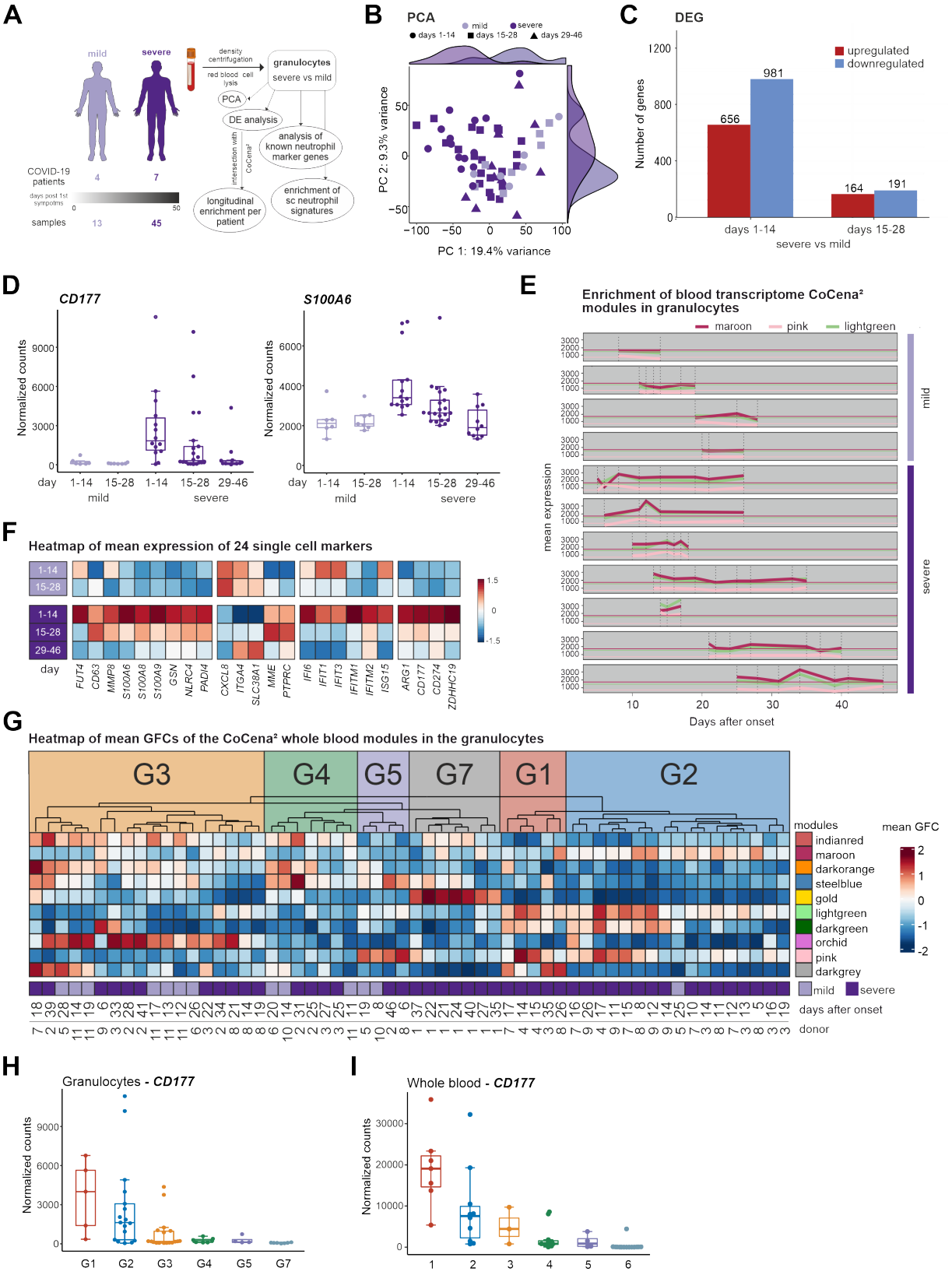
Figure S4 – related to Figure 2

(A) Heat map of mean group fold changes (GFCs) of the CoCena^2 whole blood modules in the second COVID-19 cohort for each sample. Patients are clusters by the mean GFC module expression. Severity patterns found in the whole blood CoCena^2 network were identified and patients groups were labeled accordingly (G1-G6).

(B) Agglomerative hierarchical clustering of the samples according to the GFC. Top plots present the clustering statistics (within cluster sum of squares and high average silhouette width scores) used for the generation of the six data-driven CoCena^2 sample groups G1-G6, which are plotted in the dendrogram plot.

(C) GFC heat map and hierarchical clustering for the four identified sample groups and the gene modules identified with CoCena^2 analysis.

Figure 3



1168

1169 *Figure 3 – Granulocytes from severe COVID-19 patients show a simultaneous increase in*
1170 *inflammatory and suppressive signatures*

1171 **(A)** Schema of sample processing and analysis.

1172 **(B)** PCA of all genes within the dataset mapped by COVID-19 severity status.

1173 **(C)** Bar plot of DEGs between severe and mild COVID-19 patients at day 1-14 (left) and day 15-28 (right)
1174 (FC>|2|, FDR-adj. p-value <0.05).

1175 **(D)** Boxplot of *CD177* (left) and *S100A6* (right) in mild and severe COVID-19 patients at day 1-14 and 15-
1176 28.

1177 **(E)** Mean of group fold changes (GFCs) of the modules maroon, pink and lightgreen in the granulocyte
1178 samples over time. Patients are grouped according to severity mild (top) and severe (bottom). Samples are
1179 ordered by the days after disease onset. Maroon, pink and lightgreen lines represent the overall mean of
1180 GFCs in the mild patient group.

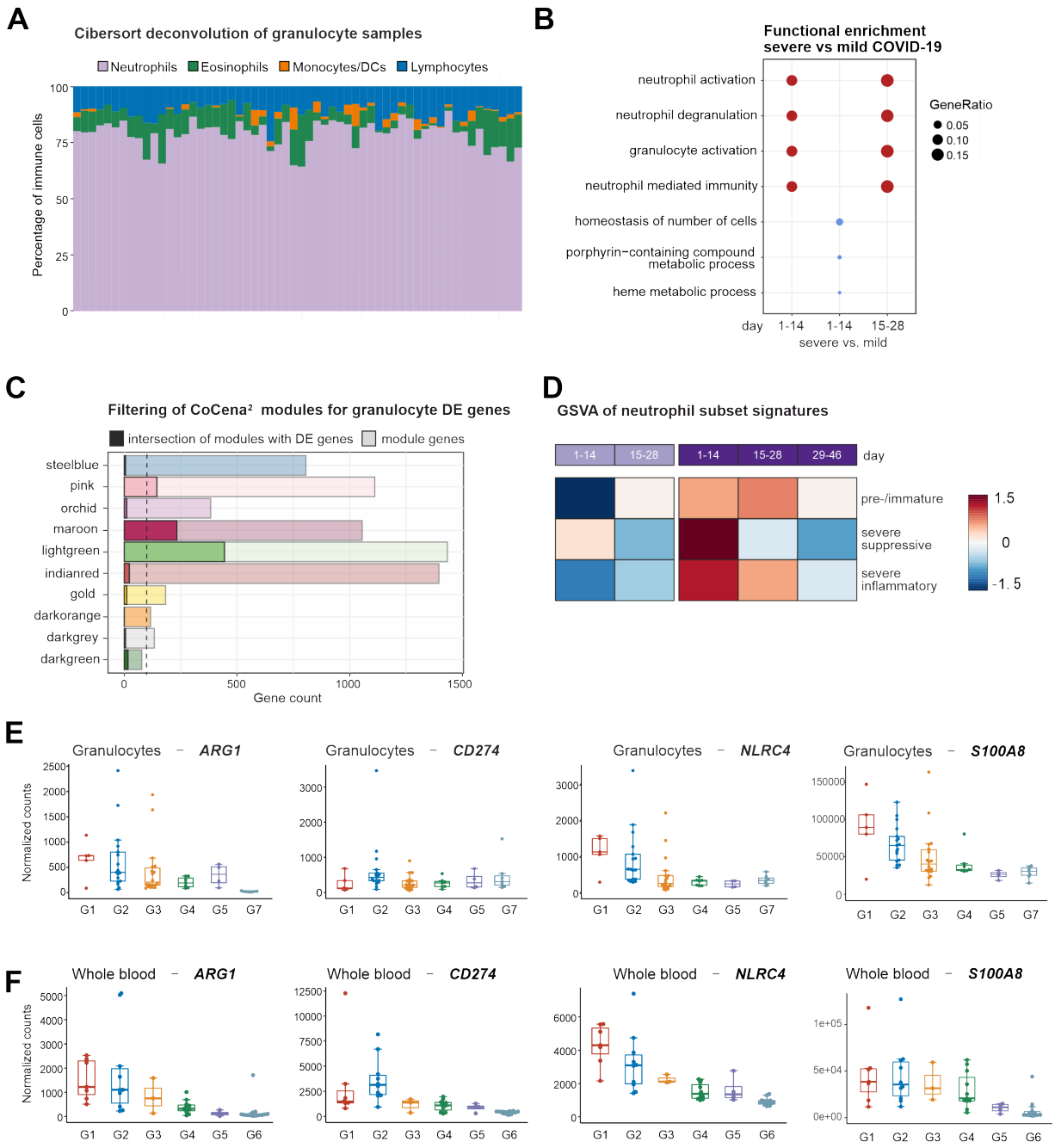
1181 **(F)** Heat map of mean expression of 24 markers in mild (top) and severe (bottom) patients ordered by days
1182 after disease onset bins (day 1-14, 15-28 and 29-46).

1183 **(G)** Heat map of mean GFCs of the CoCena[2] whole blood modules in the granulocyte samples from each
1184 individual patient. Patients are clusters by the mean GFC module expression. Severity patterns found in
1185 the whole blood CoCena[2] network were identified and patients groups were assigned accordingly (G1-G6).
1186 Patients with a distinct GFC expression pattern were labeled as G7.

1187 **(H)** Box plot of *CD177* expression in granulocytes grouped by G1-G7.

1188 **(I)** Box plot of *CD177* expression in whole blood grouped by G1-G6.

Supplemental Figure 5 - related to Figure 3



1192 *Figure S5 – related to Figure 3*

1193 (A) Cibersort computational deconvolution of the 59 granulocyte samples used in Figure 3. The LM22
1194 reference signature was used.

1195 (B) Functional enrichment analysis of the DEGs between severe and mild COVID-19 patients by GOEA.

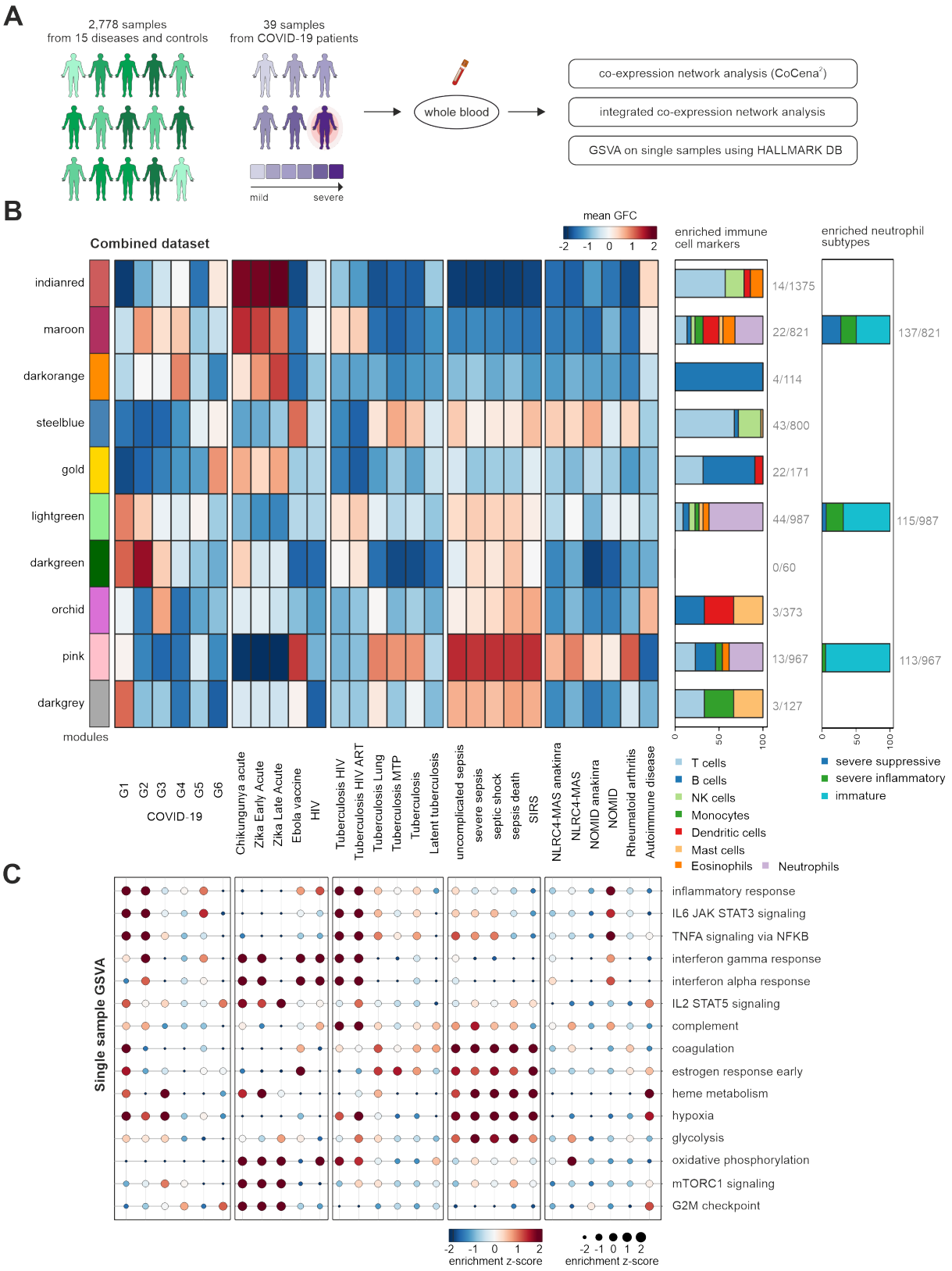
1196 (C) CoCena[2] whole blood module genes filtered for severe vs mild COVID-19 DEGs. Intersection of
1197 modules with DEGs are shown in opaque bars, filtered network module genes are shown in transparent
1198 bars.

1199 (D) GSVA of single-cell neutrophil signatures (*42*). Samples are ordered by COVID-19 severity status and
1200 days after disease onset.

1201 (E) Box plots of *ARG1*, *CD274*, *NLRC4* and *S100A6* in granulocytes grouped by G1-G7.

1202 (F) Box plot of *ARG1*, *CD274*, *NLRC4* and *S100A6* in whole blood grouped by G1-G6.

Figure 4



1203

1204 *Figure 4 – Integration with signatures from other diseases reveals COVID-19-specific*
1205 *characteristics*

1206 **(A)** Schema of analysis of the integrated dataset. The integrated dataset was analyzed with regard to
1207 expression patterns of the clusters previously identified in the whole blood COVID-19-specific co-
1208 expression network.
1209
1210 **(B)** Heat map of mean group fold changes of CoCena[2] module comparison between COVID-19 and other
1211 diseases. From left to right, the diseases are ordered by category (COVID-19, viral infections, bacterial
1212 infections and others). On the right side of the heat map, the first box plot shows the enriched immune cell
1213 markers in each module. The second box plot shows the enrichment of genes upregulated in specific
1214 neutrophil subtypes based on cross-referencing with single-cell data (*42*). Both box plots show enriched
1215 cell types in percent of total hits, absolute hits with respect to cluster size are stated aside.
1216
1217 **(C)** Gene set variation analysis was conducted for every single patient based on Hallmark gene sets as
1218 shown in Fig. 2D. The result was standardized to retrieve the z-scores, a disease mean was calculated and
1219 displayed as a dot plot with size and color correlating to the z-score. The labels on the x-axis are the same
1220 as in **(B)**.

Supplemental Figure 6 - related to Figure 4

A

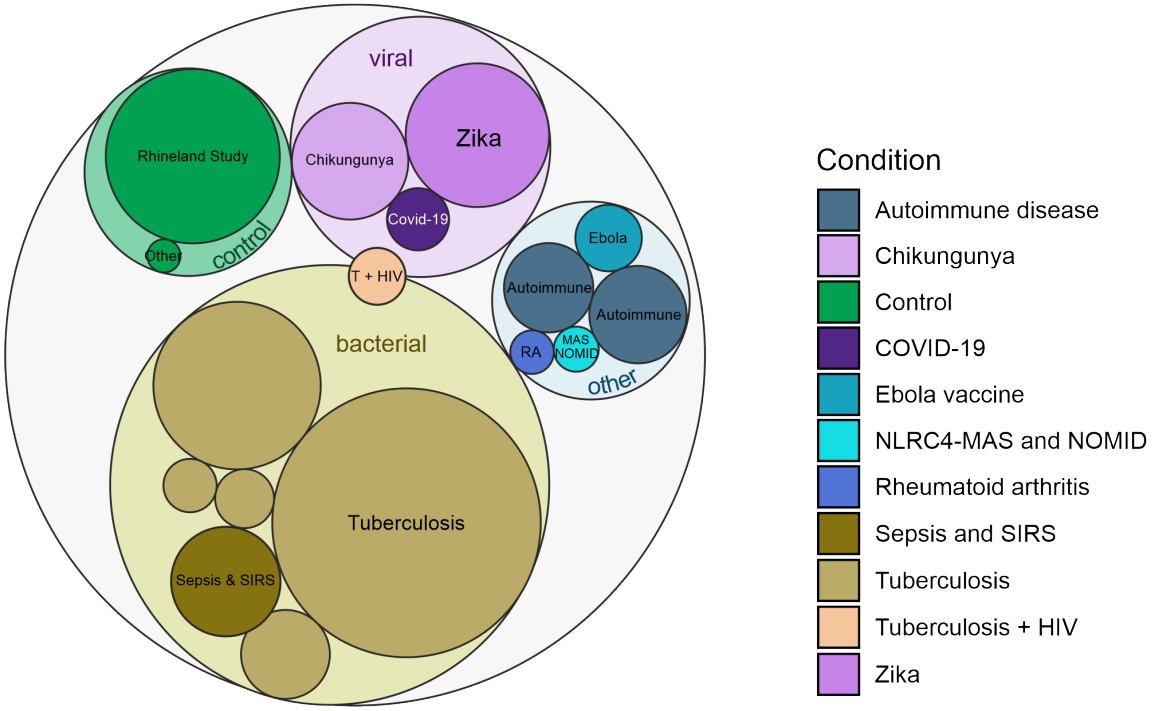
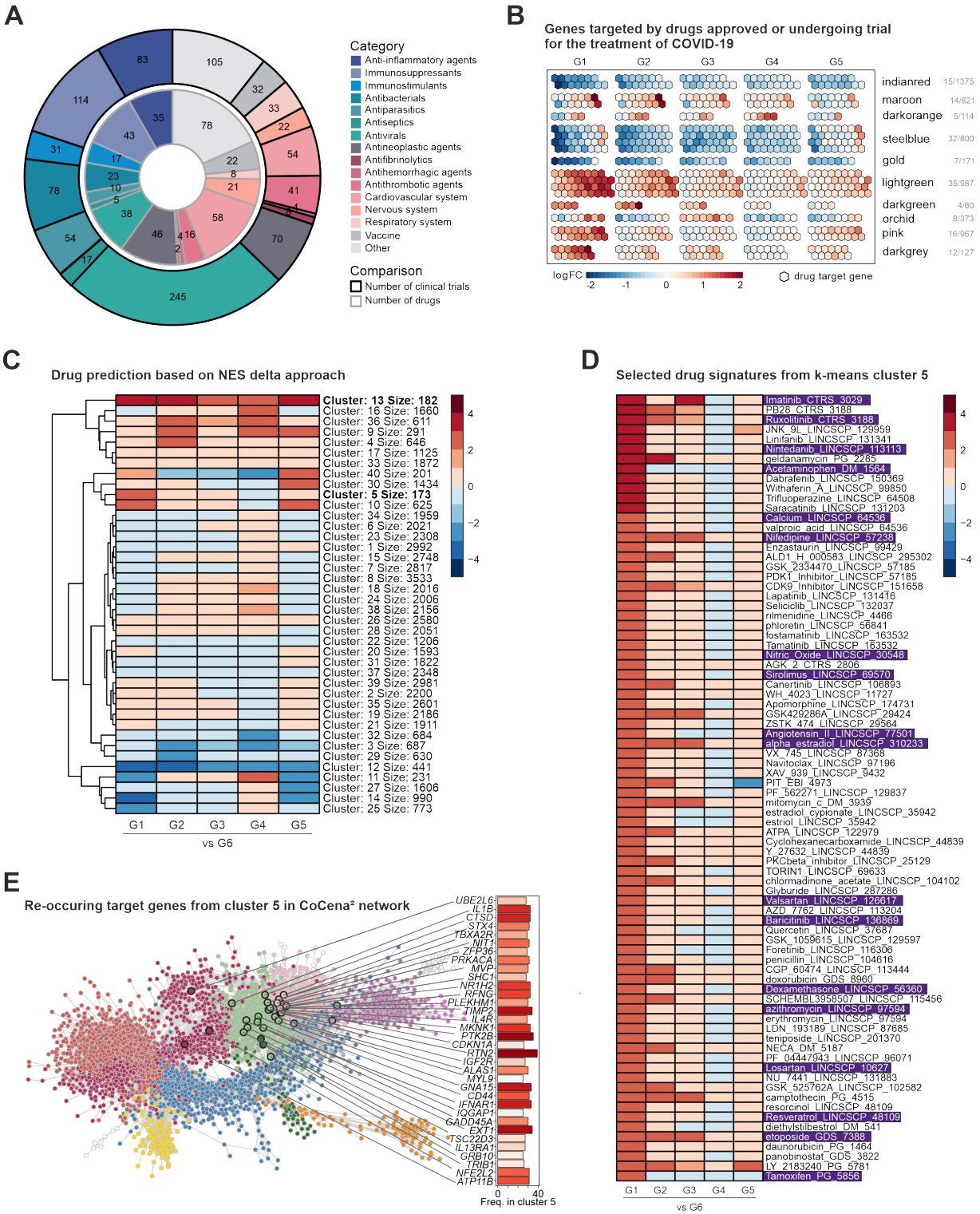


Figure S6 – related to Figure 4

(A) Overview of the composition of the integrated dataset comprising 2,817 samples: 39 COIVD-19 samples and 2,778 other conditions and controls.

Figure 5


A
Category
Anti-inflammatory agents
Immunosuppressants
Immunostimulants
Antibacterials
Antiparasitics
Antiseptics
Antivirals
Antineoplastic agents
Antifibrinolytics
Antihemorrhagic agents
Antithrombotic agents
Cardiovascular system
Nervous system
Respiratory system
Vaccine
Other
Comparison
Number of clinical trials
Number of drugs
B
Genes targeted by drugs approved or undergoing trial for the treatment of COVID-19
G1
G2
G3
G4
G5
indianred 15/1375
maroon 14/821
darkorange 5/114
steelblue 32/800
gold 7/171
lightgreen 35/987
darkgreen 4/60
orchid 8/373
pink 16/967
darkgrey 12/127
logFC
drug target gene
C
Drug prediction based on NES delta approach
Cluster: 13 Size: 182
Cluster: 16 Size: 1660
Cluster: 36 Size: 611
Cluster: 9 Size: 291
Cluster: 4 Size: 646
Cluster: 17 Size: 1125
Cluster: 33 Size: 1872
Cluster: 40 Size: 201
Cluster: 30 Size: 1434
Cluster: 5 Size: 173
Cluster: 10 Size: 625
Cluster: 34 Size: 1959
Cluster: 6 Size: 2021
Cluster: 23 Size: 2308
Cluster: 1 Size: 2992
Cluster: 15 Size: 2748
Cluster: 7 Size: 2817
Cluster: 8 Size: 3533
Cluster: 18 Size: 2016
Cluster: 24 Size: 2006
Cluster: 38 Size: 2156
Cluster: 26 Size: 2580
Cluster: 28 Size: 2051
Cluster: 22 Size: 1206
Cluster: 20 Size: 1593
Cluster: 31 Size: 1822
Cluster: 37 Size: 2348
Cluster: 39 Size: 2981
Cluster: 2 Size: 2200
Cluster: 35 Size: 2601
Cluster: 19 Size: 2186
Cluster: 21 Size: 1911
Cluster: 32 Size: 684
Cluster: 3 Size: 687
Cluster: 29 Size: 630
Cluster: 12 Size: 441
Cluster: 11 Size: 231
Cluster: 27 Size: 1606
Cluster: 14 Size: 990
Cluster: 25 Size: 773
vs G6
D
Selected drug signatures from k-means cluster 5
Imatinib_CTRS_3029
PB28_CTRS_3188
Ruxolitinib_CTRS_3188
JNK_9L_LINCSCP_129959
Linifanib_LINCSCP_131341
Nintedanib_LINCSCP_113113
geldanamycin_PG_2285
Acetaminophen_DM_1564
Dabrafenib_LINCSCP_150369
Withaferin_A_LINCSCP_99850
Trifluoperazine_LINCSCP_64508
Saracatinib_LINCSCP_131203
Calcium_LINCSCP_84536
valproic_acid_LINCSCP_64536
Nifedipine_LINCSCP_57238
Enzastaurin_LINCSCP_99429
ALD1_H_000583_LINCSCP_295302
GSK_2334470_LINCSCP_57185
PDK1_Inhibitor_LINCSCP_57185
CDK9_Inhibitor_LINCSCP_151658
Lapatinib_LINCSCP_131416
Seliciclib_LINCSCP_132037
rilmenidine_LINCSCP_4466
phloretin_LINCSCP_56841
fostamatinib_LINCSCP_163532
Tamatinib_LINCSCP_163532
Nitric_Oxide_LINCSCP_30548
AGK_2_CTRS_2806
Sirolimus_LINCSCP_69570
Ganertinib_LINCSCP_106893
WH_4023_LINCSCP_11727
Apomorphine_LINCSCP_174731
GSK429286A_LINCSCP_29424
ZSTK_474_LINCSCP_29564
Angiotensin_II_LINCSCP_77501
alpha_estradiol_LINCSCP_310233
VX_745_LINCSCP_87368
Navitoclax_LINCSCP_97196
XAV_939_LINCSCP_9432
PIT_EBI_4973
PF_562271_LINCSCP_129837
mitomycin_c_DM_3939
estradiol_cypionate_LINCSCP_35942
estriol_LINCSCP_35942
ATPA_LINCSCP_122979
Cyclohexanecarboxamide_LINCSCP_44839
Y_27632_LINCSCP_44839
PRCbeta_inhibitor_LINCSCP_25129
TORIN1_LINCSCP_69633
chlormadinone_acetate_LINCSCP_104102
Glyburide_LINCSCP_287286
Valsartan_LINCSCP_126617
AZD_7762_LINCSCP_113204
Baricitinib_LINCSCP_136869
Quercetin_LINCSCP_37687
GSK_1059615_LINCSCP_129597
Foretinib_LINCSCP_116306
penicillin_LINCSCP_104616
CGP_60474_LINCSCP_113444
doxorubicin_GDS_8960
Dexamethasone_LINCSCP_56360
SCHEMBL3958507_LINCSCP_115456
azithromycin_LINCSCP_97594
erythromycin_LINCSCP_97594
LDN_193189_LINCSCP_87685
teniposide_LINCSCP_201370
NECA_DM_5187
PF_04447943_LINCSCP_96071
Losartan_LINCSCP_10627
NU_7441_LINCSCP_131883
GSK_525762A_LINCSCP_102582
camptothecin_PG_4515
resorcinol_LINCSCP_48109
Resveratrol_LINCSCP_48109
diethylstilbestrol_DM_541
etoposide_GDS_7388
daunorubicin_PG_1464
panobinostat_GDS_3822
LY_2183240_PG_5781
Tamoxifen_PG_5856
vs G6
E
Re-occuring target genes from cluster 5 in CoCena² network
UBE2L6
IL1B
CTSD
STX4
TBXA2R
NIT1
ZFP36
PRKACA
MVP
SHC1
NR1H2
RFNG
PLEKHM1
TIMP2
IL4R
MKNK1
PTK2B
CDKN1A
RTN2
IGF2R
ALAS1
MYL9
GNA15
CD44
IFNAR1
IQGAP1
GADD45A
EXT1
TSC22D3
IL13RA1
GRB10
TRIB1
NFE2L2
ATP11B
0
40
Freq. in cluster 5

1231 *Figure 5 – Patient subgroup-specific signatures can be used to predict potential drug targets*

1232 **(A)** Overview of drugs currently used, investigated or recommended for the treatment of COVID-19 patients.
1233 The inner circle represents the number of drugs for the representative drug categories, the outer circle
1234 represents the number of clinical trials of drugs for the respective drug categories.

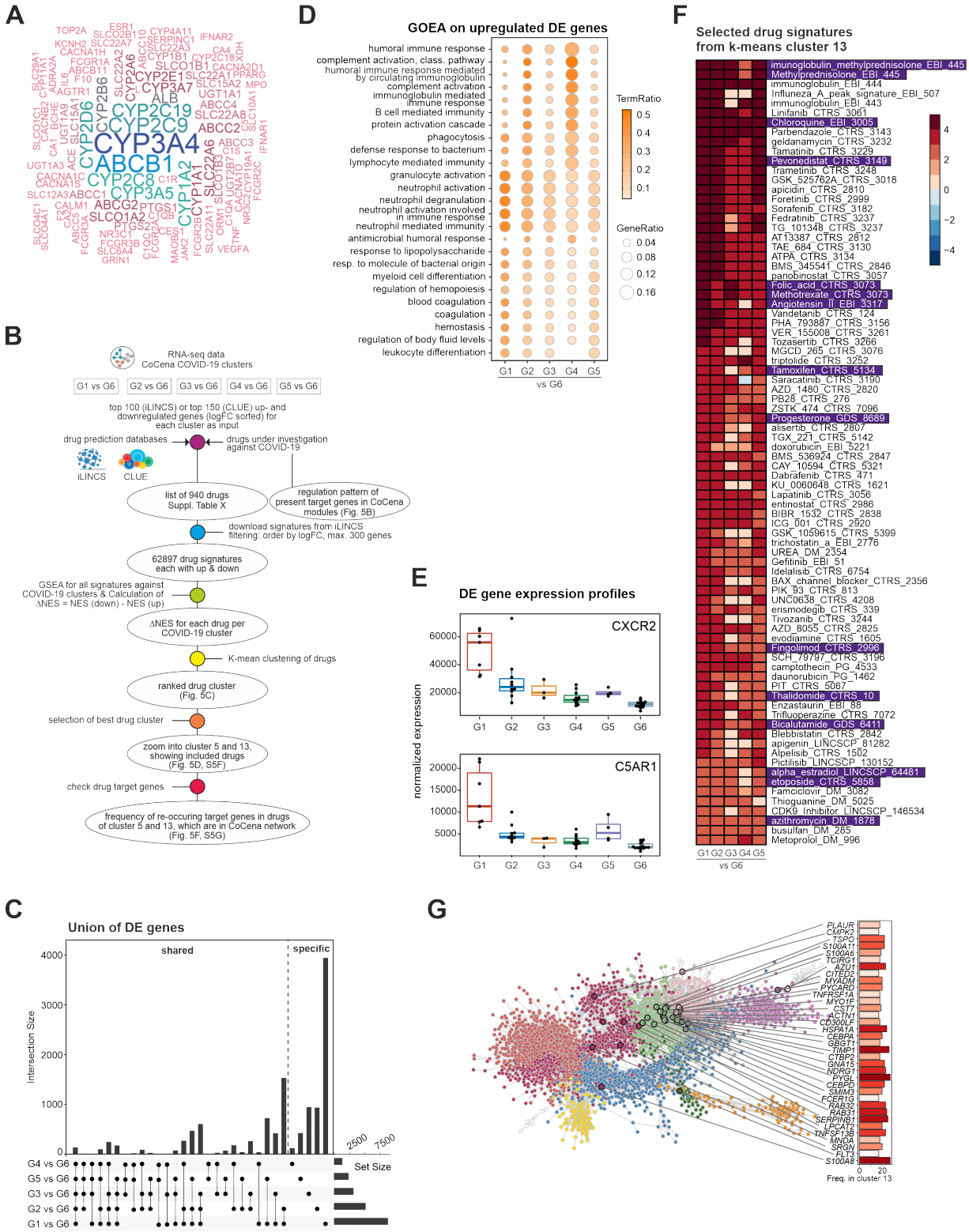
1235 **(B)** Visualization of genes targeted by drugs approved or undergoing trial for the treatment of COVID-19
1236 patients included in the whole blood co-expression network. Numbers of such genes from each module are
1237 designated on the right of the panel. Genes are represented as hexagons and colored by the expression
1238 fold change between COVID-19 patient severity group (G1-G5) and the control group (G6) (upregulated:
1239 red, downregulated: blue, not regulated: grey).

1240 **(C)** Drug predictions based on ΔNES score of drug signatures in regard to diseased patient group-specific
1241 gene expression patterns (G1-5 vs G6). Signatures were clustered by k-means clustering. A high ΔNES
1242 score accounts for drug signatures which counteract the gene expression of the patient group they are
1243 compared to. Drug signatures with a negative ΔNES score induce a gene expression pattern similar to the
1244 input. The number of signatures within a cluster determines its size.

1245 **(D)** Display of selected drug signatures from k-means cluster 5 from **(C)** showing the highest ΔNES score
1246 in the most severe COVID-19 patient group G1 and the least effect in patient group G4.

1247 **(E)** Visualization of recurring target genes in the G1 vs G6 comparison of cluster 5 signatures and their
1248 frequency mapped onto the CoCena[2] network.

Supplemental Figure 7 - related to Figure 5



1250 *Figure S7 – related to Figure 5*

1251 **(A)** Word cloud of the target genes of drugs currently investigated for the treatment of COVID-19 patients.
1252 Increasing frequency is represented by increasing size with min. frequency = 41 and max. frequency = 97.

1253 **(B)** Schematic workflow of the drug prediction analysis. Drug signatures were collected using the platforms
1254 iLINCS and CLUE. Signatures were selected by highest counteracting ΔNES score and combined with
1255 signatures of drugs under investigation from the literature.

1256 **(C)** Differentially expressed genes (FC>|2|, FDR-adj. p-value <0.05) of comparisons between groups G1-5
1257 vs G6. Vertical bar plots indicate the number of group-specific differentially expressed genes (DEGs, right)
1258 and genes shared by several groups (left), whereby the contributing groups are indicated as connected
1259 dots (bottom). Horizontal bar plots visualize the size of DEGs per group.

1260 **(D)** Gene ontology enrichment analysis of upregulated DEGs obtained for each comparison G1-G5 vs G6.
1261 Visualized are significant enrichments (adj. p-value<0.05, q-value<0.05) for the union of top 10 terms per
1262 comparison. Term ratio indicates the ratio of DEGs matching the term and the total gene number of that
1263 term.

1264 **(E)** Box plots of normalized expression of selected DEGs, upregulated in at least one comparison.

1265 **(F)** Display of selected drug signatures from k-means cluster 13 from Fig. 5C showing high ΔNES scores
1266 throughout all patient groups compared to G6.

1267 **(G)** Visualization of recurring target genes in the G1 vs G6 comparison of cluster 13 signatures and their
1268 frequency mapped onto the CoCena[2] network