EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original research article

Transcriptomic clustering of critically ill COVID-19 patients

Cecilia López-Martínez, Paula Martín-Vicente, Juan Gómez de Oña, Inés López-Alonso, Helena Gil-Peña, Elías Cuesta-Llavona, Margarita Fernández-Rodríguez, Irene Crespo, Estefanía Salgado del Riego, Raquel Rodríguez-García, Diego Parra, Javier Fernández, Javier Rodríguez-Carrio, Francisco José Jimeno-Demuth, Alberto Dávalos, Luis A Chapado, Eliecer Coto, Guillermo M Albaiceta, Laura Amado-Rodríguez

Please cite this article as: López-Martínez C, Martín-Vicente P, Gómez de Oña J, *et al*. Transcriptomic clustering of critically ill COVID-19 patients. *Eur Respir J* 2022; in press (https://doi.org/10.1183/13993003.00592-2022).

Transcriptomic clustering of critically ill COVID-19 patients

Cecilia López-Martínez[1,2,3], Paula Martín-Vicente[1,2,3], Juan Gómez de Oña[1,4,5], Inés López-Alonso[1,2,3,6], Helena Gil-Peña[1,7], Elías Cuesta-Llavona[1,4], Margarita Fernández-Rodríguez[1,3], Irene Crespo[1,2,8], Estefanía Salgado del Riego[1,9], Raquel Rodríguez-García[1,2,10], Diego Parra[1,9], Javier Fernández[1,11], Javier Rodríguez-Carrio[1,8], Francisco José Jimeno-Demuth[12], Alberto Dávalos[13], Luis A Chapado[13], Eliecer Coto[1,4,5,14], Guillermo M Albaiceta[1,2,3,7,10,*], Laura Amado-Rodríguez[1,2,3,10,14,*]

[1]Instituto de Investigación Sanitaria del Principado de Asturias. Oviedo, Spain.

[2]Centro de Investigación Biomédica en Red (CIBER)-Enfermedades respiratorias. Madrid, Spain.

[3]Instituto Universitario de Oncología del Principado de Asturias. Oviedo, Spain.

[4]Servicio de Genética Molecular. Hospital Universitario Central de Asturias. Oviedo, Spain.

[5]Red de Investigación Renal (REDINREN), Madrid, Spain.

[6]Departamento de Morfología y Biología Celular. Universidad de Oviedo. Oviedo, Spain.

[7]Servicio de Pediatría. Hospital Universitario Central de Asturias. Oviedo, Spain.

[8]Departamento de Biología Funcional. Universidad de Oviedo. Spain.

[9]Unidad de Cuidados Intensivos Polivalente. Hospital Universitario Central de Asturias. Oviedo, Spain.

[10]Unidad de Cuidados Intensivos Cardiológicos. Hospital Universitario Central de Asturias. Oviedo, Spain.

[11]Servicio de Microbiología. Hospital Universitario Central de Asturias. Oviedo, Spain.

[12]Servicio de Informática. Hospital Universitario Central de Asturias. Oviedo, Spain.

[13]Instituto Madrileño de Estudios Avanzados. (IMDEA) Alimentación, CEI UAM + CSIC, Madrid, Spain.

[14]Departamento de Medicina. Universidad de Oviedo. Oviedo, Spain.

*GMA and LAR share last authorship.

Authors for correspondence:

Guillermo M Albaiceta

Unidad de Cuidados Intensivos Cardiológicos

Hospital Universitario Central de Asturias

Avenida del Hospital Universitario s/n

33011 Oviedo. Spain.

E-mail: gma@crit-lab.org

Laura Amado-Rodríguez

Unidad de Cuidados Intensivos Cardiológicos

Hospital Universitario Central de Asturias

Avenida del Hospital Universitario s/n

33011 Oviedo. Spain.

E-mail: lar@crit-lab.org

Take-home message: Gene expression in peripheral blood determines two clusters of critically ill patients with different pathogenesis and outcomes

Abstract

Infections caused by SARS-CoV-2 may cause a severe disease, termed COVID-19, with significant mortality. Host responses to this infection, mainly in terms of systemic inflammation, have emerged as key pathogenetic mechanisms, and their modulation has shown a mortality benefit.

In a cohort of 56 critically-ill COVID-19 patients, peripheral blood transcriptomes were obtained at admission in an Intensive Care Unit (ICU) and clustered using an unsupervised algorithm. Differences in gene expression, circulating microRNAs (c-miRNA) and clinical data between clusters were assessed, and circulating cell populations estimated from sequencing data. A transcriptomic signature was defined and applied to an external cohort to validate the findings.

We identified two transcriptomic clusters characterized by expression of either interferon-related or immune checkpoint genes, respectively. Steroids have cluster-specific effects, decreasing lymphocyte activation in the former but promoting B-cell activation in the latter. These profiles have different ICU outcome, in spite of no major clinical differences at ICU admission. A transcriptomic signature was used to identify these clusters in two external validation cohorts (with 50 and 60 patients), yielding similar results.

These results reveal different underlying pathogenetic mechanisms and illustrate the potential of transcriptomics to identify patient endotypes in severe COVID-19, aimed to ultimately personalize their therapies.

Infections caused by SARS-CoV-2 have a wide range of severity, from asymptomatic to life-threatening cases. The most severe forms of Coronavirus-induced disease (termed COVID-19) [1] lead to respiratory failure fulfilling the acute respiratory distress syndrome (ARDS) criteria [2]. These critically ill patients often require mechanical ventilation and supportive therapy in an intensive care unit (ICU) and show mortality rates that range from 12 to 91% depending on patient and hospital factors [3].

Local and systemic inflammation are key pathogenetic mechanisms in severe COVID-19 [4]. Viral infection triggers a host response that involves not only anti-viral mechanisms, such as release of interferons, but may also activate a systemic, non-specific inflammatory response that has been related to multiple organ failure and death [5]. In addition to standard supportive care, the only treatments that have shown a survival benefit in critically-ill COVID-19 patients aim to modulate this inflammatory response [6]. However, it has been suggested that these treatments do not benefit patients with less severe forms of the disease or with only a mild activation of inflammation [7, 8].

There is increasing evidence that ARDS patients show different clinical features or systemic responses to severe diseases (phenotypes and endotypes respectively) [9]. Although the underlying causes responsible for this heterogeneity are not fully understood, clinical data showing different outcomes in response to a given treatment suggest that pathogenetic mechanisms may be different [10]. Therefore, identification of patient pheno/endotypes may be relevant not only for risk stratification, but also to design specific, personalized therapies in the ICU. Interestingly, whereas clustering of severe COVID-19 patients using respiratory data at ICU admission did not identify different phenotypes [11], addition of circulating biomarkers allowed the translation of the previously identified ARDS phenotypes to COVID-19 showed two groups of patients

with different responses to steroid therapy [12], highlighting the relevance of the systemic response in this setting.

Transcriptomic profiling after sequencing of whole blood RNA may be useful to identify groups of critically-ill patients with different underlying pathogenetic mechanisms [13–15]. In addition, microRNAs have been proposed to confer robustness to biological processes by reinforcing transcriptional programs, with important pathophysiological consequences [16]. Preliminary results suggest that circulating micro-RNA (c-miRNA) expression could also play a role in this setting [17]. We hypothesized that clustering of COVID-19 patients using transcriptomics at ICU admission could help to identify subgroups with different pathogenesis. To test this hypothesis, we prospectively sequenced peripheral blood RNA and serum c-miRNA at ICU admission in a cohort of COVID-19 patients, applied an unbiased clustering algorithm and compared gene expression, clinical data and outcomes in the identified subgroups. Finally, we validated our findings in an external cohort.

Methods

Study design

This prospective observational study was reviewed and approved by the regional ethics committee (Comité de Ética de la Investigación Clínica del Principado de Asturias, ref 2020.188). Informed consent was obtained from each patient's next of kin. Fifty-six consecutive patients admitted to one of the participant ICUs at Hospital Universitario Central de Asturias (Oviedo, Spain) from April to December 2020 were included in the study. Inclusion criteria were ICU admission and PCR-confirmed COVID-19. Exclusion criteria were age<18, any condition that could explain the respiratory failure other than

COVID-19, do-not-resuscitate orders or terminal status, refusal to participate or severe comorbidities that may alter the systemic response (immunosuppression, history of organ transplantation, disseminated neoplasms). All patients were managed following a standardized written clinical protocol.

Sample acquisition and processing

After inclusion, two samples of peripheral blood were drawn in the first 72 hours after ICU admission. One sample was collected in Tempus Blood RNA tubes (Thermo Fisher) to facilitate cell lysis, precipitate RNA and prevent its degradation. The other sample was immediately centrifuged to obtain serum and mixed with TRI reagent for serum RNA precipitation. These tubes were stored at -80°C until processing. Whole blood RNA was extracted by isopropanol precipitation and sequenced in an Ion S5 GeneStudio sequencer using AmpliSeq Transcriptome Human Gene Expression kits that amplify canonical human transcripts (18,574 coding genes and 2,228 non-coding genes with a complete annotation in RefSeq [https://www.ncbi.nlm.nih.gov/refseq/]). Details on RNA extraction and sequencing have been provided elsewhere [8]. FASTQ files containing RNA sequences were pseudoaligned using a reference transcriptome (http://refgenomes.databio.org) and *salmon* software [18] to obtain transcript counts.

Total serum RNA was extracted using miRNEasy kit (Qiagen), following manufacturer's instructions, and c-miRNA isolated and sequenced at BGI Genomics (Wuhan, China). c-miRNA readouts were mapped using *bowtie2* [19], with an index built using the hg38 human reference genome. Quantification of sequenced miRNAs was performed using *miRDeep2* [20] with reference human mature and hairpin miRNA sequences downloaded from miRBase (release 22, https://www.mirbase.org).

Clustering

Clustering of RNA samples was performed following a previously described protocol [21]. Briefly, log_2-transformed gene expression data (expressed as transcripts per million reads) were filtered to keep the 5% of features with the largest variance. Clusters were built based on Euclidean distances following the Ward clustering algorithm [22] and represented after dimensionality reduction using a uniform manifold approximation and projection (UMAP) algorithm [23]. Cluster p-values, indicating how strong the cluster is supported by the data (this is, the p-value with the alternative hypothesis that the cluster does not exist), were calculated by multiscale bootstrap resampling using the *pvclust* package [24] for R.

Analysis of differentially expressed genes and circulating microRNA

Gene raw counts obtained after pseudoalignment were compared between clusters using *DESeq2* [25]. Log_2 fold change for each gene between clusters and the corresponding adjusted p-value (corrected using a false discovery rate of 0.05) were calculated. Genes with an absolute log_2 fold change above 2 and an adjusted p-value lower than 0.01 were used for Gene Set Enrichment Analysis (GSEA) using the *clusterProfiler* R package [26].

A correlation analysis was performed in genes annotated to a Gene Ontology category involved in interferon pathway. Correlation coefficients between each gene pair were transformed to z-scores and the p-values for each comparison calculated using the *DGCA* package for R [27]. Genes with opposite correlations in each cluster were selected and the networks defined by their significant correlations traced.

Differentially expressed genes between clusters were also matched with the c-miRNAs expressed for each group using the *MicroRNA Target Filter* tool from *Ingenuity Pathway Analysis* (Qiagen Digital Insights), to identify predicted interactions. Intersected mRNA and miRNA datasets were filtered to explicitly pair opposed and reciprocal expression changes. Only experimentally observed predictions were considered. Key mRNA-miRNA relationships identified were overlayed onto the networks of interest to explore the predicted functionality in our datasets. Pathways related to humoral, and T and B cellular immune responses were selected as relevant. miRNAs with <3 targeted mRNAs were filtered out from the network.

Changes in gene expression after steroid therapy

To study the effects of steroids in each cluster, peripheral blood gene expression after 4 days in the ICU was assessed in 27 patients (18 assigned to CTP1 and 9 to CTP2), comparing those receiving treatment with dexamethasone (6 mg/12 h) to those who were not treated with this steroid. RNA extraction, sequencing and analysis were performed as described. Pathways with a differential response to steroids were identified after GSEA as those with significant enrichment scores with opposite signs.

Clinical data

Demographics and comorbidities were collected at ICU admission (day 1). Data on gas exchange, respiratory support, hemodynamics, received treatments and results from routine laboratory analyses were prospectively collected at days 1 and 7 after ICU admission. Patients were followed up to ICU discharge. During this period, duration of ventilatory support and vital status were collected for outcome analysis.

Circulating cell populations

Proportions of transcriptionally active circulating cells in each sample were estimated using *Immunostates* [28], a previously published deconvolution algorithm. From the original reference matrix, cell populations not commonly identified in peripheral blood (Mast cells and macrophages) were removed. Using this modified reference matrix containing expression of 318 genes for 16 different blood cell types, the percentage of each one of these types was estimated from the bulk RNAseq.

Validation

To validate our results in an external cohort, we used two publicly available dataset of 50 and 60 transcriptomes from severe and critically-ill COVID-19 patients [29, 30]. Sample acquisition was performed at enrolment. Clinical data and gene counts were downloaded from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/, accession number GSE157103)[29] or Zenodo (https://zenodo.org/record/6120249)[30]. First, we identified differentially expressed genes that best discriminate between clusters in our data, as those with an Area Under the Receiver Operating Characteristic curve (AUROC) above 0.95. A transcriptomic score was calculated as the geometric mean of these genes, and the AUROC for this score determined and a threshold between clusters was defined. Finally, raw gene expression data from validation cohorts were normalized using DESeq2, transcriptomic scores calculated, and each sample assigned to one cluster using the previously established threshold, scaled to the range of obtained values (to account for the variability in

sequencing techniques). Clinical data, outcomes and estimated cell populations (by bulk RNA-seq deconvolution as previously described) were compared between clusters.

Statistical analysis

Given the observational nature of the study and the lack of previous results, no formal sample size calculations were done. Data are expressed as median and interquartile range. Missing data were not imputed. Differences between clusters were assessed using two-tailed Wilcoxon or chi-square tests (for quantitative and qualitative data respectively). For survival analysis, patients were followed up to ICU discharge, with ICU discharge alive and spontaneously breathing being the main outcome measurement. Differences in this outcome between clusters were assessed using a competing risk model as previously described [8], and hazard ratio for the main outcome, with the corresponding 95% confidence interval, was calculated. All the analyses were performed using R v4.1.1 [31] and packages *ggplot2* [32], pROC [33] and *survival* [34], in addition to those previously cited. All the code and raw data can be found at https://github.com/Crit-Lab/COVID_clustering.

Results

Patient clustering

Peripheral gene expression was sequenced in 56 consecutive critically-ill patients (20% female, age 68 [61 - 75] year) admitted to one of the participant ICUs. Amongst 16903 genes counted, 1727 were used for hierarchical clustering (Figure 1A). The two main branches of the obtained clustering tree showed the highest p values for an alternative hypothesis that the clusters do not exist (Figure 1B and supplementary Figure 1).

Therefore, the sample was divided in two mutually exclusive groups, termed COVID-19 transcriptomic profiles (CTP) 1 and 2. Bidimensional representation of the study population using a UMAP algorithm confirmed the separation of the two clusters (Figure 1C). Supplementary figure 2 shows a heatmap with the expression of the genes used for clustering.

Differences between transcriptomic profiles

Then we assessed the overall differences in gene expression. Using an adjusted p-value cut-off point of 0.01, there were 9700 differentially expressed genes (Supplementary file 1), with 3640 having an absolute $\log_2$ fold change above 2 (Figure 2A). Interestingly, most of these genes were downregulated in CTP2. Then, GSEA was used to identify the molecular pathways involving these differentially expressed genes. One hundred and ten biological processes with significant differences between clusters were identified (Supplementary Figure 3). Among these, several categories related to the interferon-mediated response and lymphocyte activation were identified (Figure 2B), and participating genes were plotted (Figure 2C-E). Patients included in CTP1 showed an enrichment of several interferon genes, linked to the activation of a number of immune populations related to innate and adaptative responses (Figure 2C), whereas CTP2 was enriched in genes involved in B-cell receptor signaling (Figure 2D) and regulatory T-cell differentiation (Figure 2E).

In addition to these quantitative changes in expression of interferon-related genes, we explored the existence of qualitative differences between clusters. We calculated the linear correlation coefficients among the 145 genes included in the Gene Ontology categories involving interferon signaling in each cluster. There was a significant

difference between the two correlation matrices (Figure 3, $p<0.001$ calculated using a Chi-square test), thus demonstrating differences in the orchestration/structure of IFN responses between groups. In addition, pairwise differences in correlation coefficients for each gene pair were assessed. Gene pairs with correlation coefficients with an adjusted p-value for their difference below 0.05 and opposite signs in each cluster were selected, and networks including these genes traced (Figure 3 and Supplementary Figure 4). These results suggest that both clusters have a qualitatively different activation of the interferon pathway, with some genes such as HSP90AB1 and JAK1 acting as hubs with opposite correlations. Of note, CTP1 was hallmarked by strong, positive correlations among effector IFN proteins, whereas this was not the case for CTP2.

Differences in circulating cell populations

The previous results suggest that the identified clusters may have a different circulating lymphocyte profile. To further explore this finding, cell populations were estimated by deconvolution of RNAseq data. This analysis revealed a higher granulocyte proportion in patients assigned to CTP1, a lower proportion of lymphocytes and no differences in monocytes or NK cells (Figure 4A-D). Although no differences in absolute lymphocyte counts were found (645 [483 — 948] vs 730 [580 – 908] /mm3, p=0.71, Table 1), deconvolution and adjustment by total lymphocyte fraction revealed a higher proportion of CD4+ T cells (Figure 4E) and a lower proportion of CD8+ T cells (Figure 4F) and naïve B-cells (Figure 4G), with no differences in memory B-cells (Figure 4H) in this group. Detailed data on other cell populations can be found in the online supplement (supplementary figure 5).

Potential regulatory miRNAs

To identify c-miRNA potentially related to the observed changes in RNA expression and immune cell populations, we analyzed miRNA content using the *MiRNA Target Filter* tool included in *Ingenuity Pathway Analysis*. After filtering by experimentally confirmed miRNA-gene relationships, and only opposed changes in miRNA/gene expression levels, 83 miRNAs targeting 608 genes were identified in our dataset of differentially expressed genes. Given the observed differences in lymphocyte populations, we focused on miRNAs involved in humoral and cellular immune regulation (29 miRNAs and 151 genes). Paired miRNA-gene networks are depicted in Supplementary Figure 6 (104 downregulated genes/18 predicted upregulated miRNAs) and Figure 5 (47 upregulated genes/11 predicted downregulated miRNAs), with an overlay including differentially expressed genes between CTP1 and CTP2. miRNAs predicted to regulate expression of these genes were identified and compared (Figure 5B-H). Among these, counts of miR-145a-5p and miR-181-5p were significatively lower in CTP2 (Figure 5C and 5D respectively).

Cluster-specific effects of steroids

To assess cluster-specific effects of steroids, we compared gene expression after 4 days of ICU stay in patients with and without steroids in each cluster. Although steroids modified the transcriptomic profile in both clusters, the overlap in differentially expressed genes between clusters was minimal (Figures 6A-B). When pathways with divergent responses were assessed (Figure 6C), we found that steroids downregulated T- and B- cell activation and IL production and activated JAK-STAT signaling only in

patients from the CTP1 cluster. In opposite, steroid therapy was related to B-cell activation in patients assigned to CTP2.

Clinical differences and outcome

Clinical differences between clusters at ICU admission were studied (Table 1). There were no significant differences in demographic and clinical variables other than a higher neutrophil count in cluster CTP1, with no differences in lymphocyte counts. Patients assigned to CTP2 cluster showed more ventilator-free days during the first 28 days in ICU (Table 1). In the survival analysis, after adjusting for age, sex, and need for intubation during the ICU stay, assignation to CTP2 increased the probability of ICU discharge alive and spontaneously breathing (HR 2.00 [1.08 – 3.70], p=0.028, Figure 7). Other used biomarkers as neutrophil count, neutrophil-to-lymphocyte ratio or C-reactive protein showed only a moderate performance for cluster assignment (AUROCs of 0.74 [0.61 – 0.88], 0.73 [0.57 – 0.89] and 0.53 [0.31 – 0.77], respectively). Replacing cluster assignment with neutrophil-to-lymphocyte ratio or C-reactive protein did not yield a statistically significant HR in the survival analyses (HR 0.958 [0.908 – 1.011], p=0.122 for neutrophil-to-lymphocyte ratio; HR 0.986 [0.957 – 1.016], p=0.365 for C-reactive protein).

Definition of a transcriptomic signature and external validation

To apply our findings to an external cohort, we first developed a characteristic gene signature that allows assignation to one cluster using gene expression data. We focused on genes upregulated in CTP2, as they constitute a relatively small group, given the massive gene downregulation in this group. Among these 117 upregulated genes, 15

(*BCL2, CARD11, CD247, CD7, CD81, CLSTN1, E2F6, MCM5, PARP1, PNPO, RASGRP1, RCC2, RPTOR, RUNX3* and *ZAP70*) had an AUROC to identify CTP2 higher than 0.95. Expression of these genes was synthesized into a transcriptomic score. As expected, the score was higher in CTP2 (Supplementary figure 7A), with an AUROC of 0.99 (95% CI 0.97 – 1) (Supplementary figure 7B). In a Cox-regression analysis including this transcriptomic score, age, sex and need for mechanical ventilation, the score was correlated to ICU discharge (HR 1.202 [1.041 – 1.387] per 100 points increase in transcriptomic score, p=0.012, supplemental figure 8). Based on these results, a cut-off point of 250 in this score, aimed to include all CTP2 cases, was chosen.

Then, this transcriptomic score was calculated in two external cohorts. Regarding the first cohort (n=50), 13 patients were classified as CTP1 and 37 as CTP2. Comparisons between these clusters are shown in Table 2. In spite of no significant differences in age, sex, APACHE-II or SOFA scores, patients assigned to CTP2 showed more ventilator-free days at day 28 of ICU stay, and the percentage of patients with zero ventilator-free days at day 28 was lower in CTP2. In the second validation cohort (n=60), 22 patients were classified as CTP1 and 38 as CTP2, after rescaling the cut-off point to account for differences in sequencing technology and depth. Resembling the previous results, there were no differences in age and sex, but mortality by day 28 was higher in patients assigned to CTP1 (Table 2). Deconvolution of peripheral blood transcriptomes in both validation cohorts recapitulated some of the differences observed in the discovery cohort, including higher neutrophil counts and lower proportions of CD8+ T-cells in CTP1 (Supplementary Figures 9 and 10).

Discussion

Our results show that unsupervised transcriptomic clustering of critically ill COVID-19 patients at ICU admission, results in two groups with different immune profiles, response to steroids and outcome. Application of a cluster-specific score to two independent cohorts confirmed this result. These findings suggest there are specific COVID-19 endotypes with different underlying immunopathogenesis and outcomes.

Clustering strategies have been proposed to identify different subgroups of critically ill patients with respiratory failure that may help to personalize treatments. In ARDS, a hyperinflammatory/reactive phenotype [9, 35], characterized by markers of acute inflammation and tissue hypoxia, has been linked to higher mortality rates and could specifically benefit from fluid restriction, higher PEEP levels or protective ventilation [36], in contraposition to the uninflamed phenotype. Of note, causes of ARDS were different between phenotypes, with a higher incidence of sepsis in the hyperinflamed/reactive group. Clustering of COVID-19 patients using respiratory data failed to identify phenotypes at ICU admission [11]. Addition of clinically available biomarkers allowed a direct translation of the inflammatory/reactive framework in two cohorts [12, 37]. Focusing on a single disease (COVID-19) rather than a syndrome (ARDS) could result in a reduced phenotypic variability, thus increasing the informative value of the systemic response evaluated by circulating biomarkers. Clinically available markers as neutrophil-to-lymphocyte ratio or C-reactive protein are usually elevated in severe forms of COVID-19 [38], but their role to stratify patients is yet to be determined and, in our study, failed to predict outcome or cluster assignment.

In this setting, transcriptomic clustering may offer several advantages by including a large number of features for classification, reduced intervention times and absence of imputed or not available data, although the superiority of this approach remains to be

demonstrated. Increasing evidence points at c-miRNA as biomarkers with pathogenetic implications given their role as modulators of gene expression. Point-of-care devices under development would allow the quantification of our 15-gene signature or validated cluster-specific c-miRNA at the bedside, to rapidly identify these patient endotypes and predict outcomes [39].

Bulk peripheral blood RNAseq has been used to study COVID-19 pathogenesis, by comparing cases with different severity or against healthy controls [40–42]. Our approach included only severe cases, revealing two different clusters that include quantitative and qualitative differences in the regulation of the immune response to SARS-CoV-2 infection and different responses to steroids. Of note, these biological disparities occur despite no differences in clinical variables, suggesting that clusters reflect endotypes with specific pathogenetic mechanisms and may outperform clinical diagnostic instruments.

CTP1 is characterized by an interferon-driven response and CD4+ T lymphocyte activation, that have been linked to a worse outcome [43, 44]. miR-145a-5p and miR-181-5p, which play key roles promoting granulopoiesis [45] and CD4+ T-cell maturation [16, 46] respectively, were upregulated in this cluster. Steroids downregulated genes involved in lymphocyte activation, but upregulated the JAK/STAT pathway in CTP1, which can promote further overexpression of miR-181 family members [47, 48]. JAK/STAT pathway can be activated by IL-6 and has been related to mortality in COVID-19 patients [49, 50].

CTP2 cluster, with better outcome, is characterized by B-cell and Treg activation and upregulation of immune checkpoints such as BCL2 and immunoglobulin and TNF superfamilies. Of note, BCL2 and TNF superfamilies are targeted by miR-181, which is

decreased in this cluster and has been described to induce immunoparalysis and block immune checkpoints [51]. Dysregulation of other immune checkpoints has been also linked to mortality in COVID-19 patients [52]. In this group, steroids further promoted B-cell activation. Collectively, these results raise the hypothesis that steroids may help to further regulate the inflammatory response in CTP2, but activate JAK/STAT- dependent immune response in CTP1, which in turn could partly explain our differences in ICU outcome and the proposed synergic effects of steroids and IL-6 / JAK blockade in COVID-19 [53].

Our results have several limitations. First, the sample size is reduced, so we cannot exclude the existence of additional clusters with other underlying pathogenetic mechanisms, or that different clustering parameters or strategies may yield different results. However, unbiased p-values associated to the identified clusters were high, and the results confirmed in two independent validation cohorts. It must be noted that time of sampling differs among cohorts (first 72 hours after ICU admission in our study and in the COMBAT cohort [30], between days 1 and 6 in the study by Overmyer et al. [29]). We do not have data to define specific time windows. However, the consistency of the results among studies reinforces the external validity of our clustering strategy. Third, cell populations were estimated by deconvolution of the bulk transcriptome and should be confirmed using single cell RNA seq or flow cytometry. Finally, although our data show different effects of steroids in each cluster, it is unclear if therapeutic immunomodulation may impact outcomes in a cluster-specific manner.

In summary, our results show that transcriptomic clustering using peripheral blood RNA at ICU admission allows the identification of two groups of critically-ill COVID-19 with different immune profile and outcome. These findings could be useful for risk

stratification of these patients and help to identify specific profiles that could benefit from personalized treatments aimed to modulate the inflammatory response or its consequences.

Acknowledgements

The authors want to thank all the personnel at the participating ICUs and laboratories for their support during the development of the study. This work is supported by Centro de Investigación Biomédica en Red (CIBER)-Enfermedades Respiratorias (CB17/06/00021), Instituto de Salud Carlos III (grants PI20/01360 and PI21/01592, FEDER funds) and Fundació La Marató de TV3 (413/C/2021). CLM is supported by Ministerio de Universidades, Spain (FPU18/02965). RRG is supported by a grant from Instituto de Salud Carlos III (CM20/0083). PMV is supported by a grant from Instituto de Salud Carlos III (FI21/00168). Instituto Universitario de Oncología del Principado de Asturias is supported by Fundación Liberbank. AD is supported by the Comunidad de Madrid and European Regional Development Fund (REACT EU Program "FACINGLCOVID-CM").

Author contributions

Study design and supervision: GMA, LAR. Sample acquisition and processing: MFR. RNA Sequencing: CLM, PMV, ILA, JGdO, HGP, ECL, IC, EC. MiRNA sequencing: CLM, AD, LACh. Patient inclusion, follow-up, and data collection: EsdR, RRG, DP, FJJD, GMA, LAR. Data analysis: CLM, PMV, LACh, GMA, LAR. Results discussion: CLM, PMV, ILA, IC, JRC, AD, EC, GMA, LAR. Manuscript writing: CLM, GMA, LAR. Manuscript review and editing: All authors.

Competing interests.

The authors declare no competing interests.

Data availability.

All data and code used in the manuscript have been deposited in Git-Hub (https://github.com/Crit-Lab/COVID_clustering). FASTQ files with RNA and miRNA reads have been deposited in Gene Expression Omnibus (accession number GSE197259, available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197259).

Corresponding authors

Correspondence and requests to Laura Amado-Rodríguez (lar@crit-lab.org) or Guillermo M Albaiceta (gma@crit-lab.org).

References

1. Grasselli G, Greco M, Zanella A, Albano G, Antonelli M, Bellani G, Bonanomi E, Cabrini L, Carlesso E, Castelli G, Cattaneo S, Cereda D, Colombo S, Coluccello A, Crescini G, Forastieri Molinari A, Foti G, Fumagalli R, Iotti GA, Langer T, Latronico N, Lorini FL, Mojoli F, Natalini G, Pessina CM, Ranieri VM, Rech R, Scudeller L, Rosano A, Storti E, et al. Risk Factors Associated With Mortality Among Patients With COVID-19 in Intensive Care Units in Lombardy, Italy. *JAMA Intern. Med.* 2020; 180: 1345–1355.

2. Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, Camporota L, Slutsky AS. Acute respiratory distress syndrome: the Berlin Definition. *JAMA* 2012; 307: 2526–2533.

3. Churpek MM, Gupta S, Spicer AB, Parker WF, Fahrenbach J, Brenner SK, Leaf DE, STOP-COVID Investigators. Hospital-Level Variation in Death for Critically Ill Patients with COVID-19. *Am. J. Respir. Crit. Care Med.* 2021; 204.

4. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* 2020; 20: 363–374.

5. Wong L-YR, Perlman S. Immune dysregulation and immunopathology induced by SARS-CoV-2 and related coronaviruses - are we our own worst enemy? *Nat. Rev. Immunol.* 2022; 22: 47–56.

6. Clinical management of COVID-19 patients: living guideline, 23 November 2021 [Internet]. [cited 2022 Jan 25].Available from: https://app.magicapp.org/#/guideline/j1WBYn.

7. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, Staplin N, Brightling C, Ustianowski A, Elmahi E, Prudon B, Green C, Felton T, Chadwick D, Rege K, Fegan C, Chappell LC, Faust SN, Jaki T, Jeffery K, Montgomery A, Rowan K, Juszczak E, Baillie JK, Haynes R, Landray MJ. Dexamethasone in Hospitalized Patients with Covid-19. *N. Engl. J. Med.* 2021; 384: 693–704.

8. Amado-Rodríguez L, Salgado Del Riego E, Gomez de Ona J, López Alonso I, Gil-Pena H, López-Martínez C, Martín-Vicente P, Lopez-Vazquez A, Gonzalez Lopez A, Cuesta-Llavona E, Rodriguez-Garcia R, Boga JA, Elena Alvarez-Arguelles M, Mayordomo-Colunga J, Vidal-Castineira JR, Crespo I, Fernandez M, Criado L, Salvadores V, Jimeno-Demuth FJ, Blanch L, Prieto B, Fernandez-Fernandez A, Lopez-Larrea C, Coto E, Albaiceta GM. Effects of IFIH1 rs1990760 variants on systemic inflammation and outcome in critically ill COVID-19 patients in an observational translational study. *eLife* 2022; 11: e73012.

9. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, Network NA. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014; 2: 611–620.

10. Calfee CS, Delucchi KL, Sinha P, Matthay MA, Hackett J, Shankar-Hari M, McDowell C, Laffey JG, O'Kane CM, McAuley DF, Irish Critical Care Trials Group. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir. Med.* 2018; 6: 691–698.

11. Bos LDJ, Sjoding M, Sinha P, Bhavani SV, Lyons PG, Bewley AF, Botta M, Tsonas AM, Serpa Neto A, Schultz MJ, Dickson RP, Paulus F, PRoVENT-COVID collaborative group. Longitudinal respiratory subphenotypes in patients with COVID-19-related acute respiratory distress syndrome: results from three observational cohorts. *Lancet Respir. Med.* 2021; 9: 1377–1386.

12. Sinha P, Furfaro D, Cummings MJ, Abrams D, Delucchi K, Maddali MV, He J, Thompson A, Murn M, Fountain J, Rosen A, Robbins-Juarez SY, Adan MA, Satish T, Madhavan M, Gupta A, Lyashchenko AK, Agerstrand C, Yip NH, Burkart KM, Beitler JR, Baldwin MR, Calfee CS, Brodie D, O'Donnell MR. Latent Class Analysis Reveals COVID-19-related Acute Respiratory Distress Syndrome Subgroups with Differential Responses to Corticosteroids. *Am. J. Respir. Crit. Care Med.* 2021; 204: 1274–1285.

13. Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, Nürnberg P, Schultz MJ, Horn J, Cremer OL, Bonten MJ, Hinds CJ, Wong HR, Knight JC, van der Poll T, MARS consortium. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *Lancet Respir. Med.* 2017; 5: 816–826.

14. Baghela A, Pena OM, Lee AH, Baquir B, Falsafi R, An A, Farmer SW, Hurlburt A, Mondragon-Cardona A, Rivera JD, Baker A, Trahtemberg U, Shojaei M, Jimenez-Canizales CE, Dos Santos CC, Tang B, Bouma HR, Cohen Freue GV, Hancock REW. Predicting sepsis severity at first clinical presentation: The role of endotypes and mechanistic signatures. *EBioMedicine* 2022; 75: 103776.

15. Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, Rautanen A, Gordon AC, Garrard C, Hill AVS, Hinds CJ, Knight JC. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. *Lancet Respir. Med.* 2016; 4: 259–271.

16. Ebert MS, Sharp PA. Roles for microRNAs in conferring robustness to biological processes. *Cell* 2012; 149: 515–524.

17. de Gonzalo-Calvo D, Benítez ID, Pinilla L, Carratalá A, Moncusí-Moix A, Gort-Paniello C, Molinero M, González J, Torres G, Bernal M, Pico S, Almansa R, Jorge N, Ortega A, Bustamante-Munguira E, Gómez JM, González-Rivera M, Micheloud D, Ryan P, Martinez A, Tamayo L, Aldecoa C, Ferrer R, Ceccato A, Fernández-Barat L, Motos A, Riera J, Menéndez R, Garcia-Gasulla D, Peñuelas O, et al. Circulating microRNA profiles predict the severity of COVID-19 in hospitalized patients. *Transl. Res. J. Lab. Clin. Med.* 2021; 236: 147–159.

18. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 2017; 14: 417–419.

19. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 2012; 9: 357–359.

20. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 2012; 40: 37–52.

21. Jaskowiak PA, Costa IG, Campello RJGB. Clustering of RNA-Seq samples: Comparison study on cancer data. *Methods San Diego Calif* 2018; 132: 42–49.

22. Ward JH. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* Taylor & Francis; 1963; 58: 236–244.

23. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* The Open Journal; 2018; 3: 861.

24. Suzuki R, Terada Y, Shimodaira H. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling [Internet]. 2019.Available from: https://CRAN.R-project.org/package=pvclust.

25. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 550.

26. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. N. Y. N* 2021; 2: 100141.

27. McKenzie AT, Katsyv I, Song W-M, Wang M, Zhang B. DGCA: A comprehensive R package for Differential Gene Correlation Analysis. *BMC Syst. Biol.* 2016; 10: 106.

28. Vallania F, Tam A, Lofgren S, Schaffert S, Azad TD, Bongen E, Haynes W, Alsup M, Alonso M, Davis M, Engleman E, Khatri P. Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. *Nat. Commun.* 2018; 9: 4735.

29. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, Meyer JG, Quan Q, Muehlbauer LK, Trujillo EA, He Y, Chopra A, Chieng HC, Tiwari A, Judson MA, Paulson B, Brademan DR, Zhu Y, Serrano LR, Linke V, Drake LA, Adam AP, Schwartz BS, Singer HA, Swanson S, Mosher DF, Stewart R, Coon JJ, Jaitovich A. Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Syst.* 2021; 12: 23-40.e7.

30. COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell* 2022; 185: 916-938.e58.

31. R Core Team. R: A language and environment for statistical computing [Internet]. [cited 2014 Jun 10].Available from: http://www.R-project.org/.

32. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016.Available from: https://ggplot2.tidyverse.org.

33. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77.

34. Therneau TM. A Package for Survival Analysis in R [Internet]. 2020.Available from: https://CRAN.R-project.org/package=survival.

35. Bos LDJ, Scicluna BP, Ong DSY, Cremer O, van der Poll T, Schultz MJ. Understanding Heterogeneity in Biologic Phenotypes of Acute Respiratory Distress Syndrome by Leukocyte Expression Profiles. *Am. J. Respir. Crit. Care Med.* 2019; 200: 42–50.

36. Amado-Rodríguez L, Del Busto C, López-Alonso I, Parra D, Mayordomo-Colunga J, Arias-Guillén M, Albillos-Almaraz R, Martín-Vicente P, López-Martínez C, Huidobro C, Camporota L, Slutsky AS, Albaiceta GM. Biotrauma during ultra-low tidal volume ventilation and venoarterial extracorporeal membrane oxygenation in cardiogenic shock: a randomized crossover clinical trial. *Ann. Intensive Care* 2021; 11: 132.

37. Ranjeva S, Pinciroli R, Hodell E, Mueller A, Hardin CC, Thompson BT, Berra L. Identifying clinical and biochemical phenotypes in acute respiratory distress syndrome secondary to coronavirus disease-2019. *EClinicalMedicine* 2021; 34: 100829.

38. Sayah W, Berkane I, Guermache I, Sabri M, Lakhal FZ, Yasmine Rahali S, Djidjeli A, Lamara Mahammed L, Merah F, Belaid B, Berkani L, Lazli NZ, Kheddouci L, Kadi A, Ouali M, Khellafi R, Mekideche D, Kheliouen A, Hamidi RM, Ayoub S, Raaf NB, Derrar F, Gharnaout M, Allam I, Djidjik R. Interleukin-6, procalcitonin and neutrophil-to-lymphocyte ratio: Potential immune-inflammatory parameters to identify severe and fatal forms of COVID-19. *Cytokine* 2021; 141: 155428.

39. Brakenridge SC, Starostik P, Ghita G, Midic U, Darden D, Fenner B, Wacker J, Efron PA, Liesenfeld O, Sweeney TE, Moldawer LL. A Transcriptomic Severity Metric That Predicts Clinical Outcomes in Critically Ill Surgical Sepsis Patients. *Crit. Care Explor.* 2021; 3: e0554.

40. Vanderbeke L, Van Mol P, Van Herck Y, De Smet F, Humblet-Baron S, Martinod K, Antoranz A, Arijs I, Boeckx B, Bosisio FM, Casaer M, Dauwe D, De Wever W, Dooms C, Dreesen E, Emmaneel A, Filtjens J, Gouwy M, Gunst J, Hermans G, Jansen S, Lagrou K, Liston A, Lorent N, Meersseman P, Mercier T, Neyts J, Odent J, Panovska D, Penttila PA, et al. Monocyte-driven atypical cytokine storm and aberrant neutrophil activation as key mediators of COVID-19 disease severity. *Nat. Commun.* 2021; 12: 4117.

41. Bost P, De Sanctis F, Canè S, Ugel S, Donadello K, Castellucci M, Eyal D, Fiore A, Anselmi C, Barouni RM, Trovato R, Caligola S, Lamolinara A, Iezzi M, Facciotti F, Mazzariol A, Gibellini D, De Nardo P, Tacconelli E, Gottin L, Polati E, Schwikowski B, Amit I, Bronte V. Deciphering the state of immune silence in fatal COVID-19 patients. *Nat. Commun.* 2021; 12: 1428.

42. Chan Y-H, Fong S-W, Poh C-M, Carissimo G, Yeo NK-W, Amrun SN, Goh YS, Lim J, Xu W, Chee RS-L, Torres-Ruesta A, Lee CY-P, Tay MZ, Chang ZW, Lee W-H, Wang B, Tan S-Y, Kalimuddin S, Young BE, Leo Y-S, Wang C-I, Lee B, Rötzschke O, Lye DC, Renia L, Ng LFP. Asymptomatic COVID-19: disease tolerance with efficient anti-viral immunity against SARS-CoV-2. *EMBO Mol. Med.* 2021; 13: e14045.

43. Pallotto C, Suardi LR, Esperti S, Tarquini R, Grifoni E, Meini S, Valoriani A, Di Martino S, Cei F, Sisti E, Piani F, Botta A, Salomoni E, Baragli F, Blanc P. Increased CD4/CD8 ratio as a risk factor for critical illness in coronavirus disease 2019 (COVID-19): a retrospective multicentre study. *Infect. Dis. Lond. Engl.* 2020; 52: 675–677.

44. Galani I-E, Rovina N, Lampropoulou V, Triantafyllia V, Manioudaki M, Pavlos E, Koukaki E, Fragkou PC, Panou V, Rapti V, Koltsida O, Mentis A, Koulouris N, Tsiodras S, Koutsoukou A, Andreakos E. Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. *Nat. Immunol.* 2021; 22: 32–40.

45. Allantaz F, Cheng DT, Bergauer T, Ravindran P, Rossier MF, Ebeling M, Badi L, Reis B, Bitter H, D'Asaro M, Chiappe A, Sridhar S, Pacheco GD, Burczynski ME, Hochstrasser D, Vonderscher J, Matthes T. Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PloS One* 2012; 7: e29979.

46. Li Q-J, Chau J, Ebert PJR, Sylvester G, Min H, Liu G, Braich R, Manoharan M, Soutschek J, Skare P, Klein LO, Davis MM, Chen C-Z. miR-181a is an intrinsic modulator of T cell sensitivity and selection. *Cell* 2007; 129: 147–161.

47. Chen S-L, Cai G-X, Ding H-G, Liu X-Q, Wang Z-H, Jing Y-W, Han Y-L, Jiang W-Q, Wen M-Y. JAK/STAT signaling pathway-mediated microRNA-181b promoted blood-brain barrier impairment by targeting sphingosine-1-phosphate receptor 1 in septic rats. *Ann. Transl. Med.* 2020; 8: 1458.

48. Assmann JLJC, Leon LG, Stavast CJ, van den Bogaerdt SE, Schilperoord-Vermeulen J, Sandberg Y, Bellido M, Erkeland SJ, Feith DJ, Loughran TP, Langerak AW. miR-181a is a novel player in the STAT3-mediated survival network of TCRαβ+ CD8+ T large granular lymphocyte leukemia. *Leukemia* 2022; 36: 983–993.

49. Spinelli FR, Conti F, Gadina M. HiJAKing SARS-CoV-2? The potential role of JAK inhibitors in the management of COVID-19. *Sci. Immunol.* 2020; 5: eabc5367.

50. Kalil AC, Patterson TF, Mehta AK, Tomashek KM, Wolfe CR, Ghazaryan V, Marconi VC, Ruiz-Palacios GM, Hsieh L, Kline S, Tapson V, Iovine NM, Jain MK, Sweeney DA,

El Sahly HM, Branche AR, Regalado Pineda J, Lye DC, Sandkovsky U, Luetkemeyer AF, Cohen SH, Finberg RW, Jackson PEH, Taiwo B, Paules CI, Arguinchona H, Erdmann N, Ahuja N, Frank M, Oh M-D, et al. Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. *N. Engl. J. Med.* 2021; 384: 795–807.

51. Dan C, Jinjun B, Zi-Chun H, Lin M, Wei C, Xu Z, Ri Z, Shun C, Wen-Zhu S, Qing-Cai J, Wu Y. Modulation of TNF-α mRNA stability by human antigen R and miR181s in sepsis-induced immunoparalysis. *EMBO Mol. Med.* 2015; 7: 140–157.

52. Avendaño-Ortiz J, Lozano-Rodríguez R, Martín-Quirós A, Terrón V, Maroun-Eid C, Montalbán-Hernández K, Valentín-Quiroga J, García-Garrido MÁ, Del Val EM, Del Balzo-Castillo Á, Peinado M, Gómez L, Herrero-Benito C, Rubio C, Casalvilla-Dueñas JC, Gómez-Campelo P, Pascual-Iglesias A, Del Fresno C, Aguirre LA, López-Collazo E. The immune checkpoints storm in COVID-19: Role as severity markers at emergency department admission. *Clin. Transl. Med.* 2021; 11: e573.

53. Stebbing J, Lauschke VM. JAK Inhibitors - More Than Just Glucocorticoids. *N. Engl. J. Med.* 2021; 385: 463–465.

Figure legends

Figure 1. Patient clustering. A: Clustering strategy based on peripheral blood RNAseq, using the 5% genes with the highest variance among samples. B: Hierarchical clustering tree, showing the p-values (corresponding to the alternative hypothesis that the cluster does not exist) of the two main clusters. C: Uniform manifold approximation and projection (UMAP) showing a bidimensional representation of all the samples and clusters. TPM: Transcripts per million reads.

Figure 2. Differentially expressed genes between COVID-19 transcriptomic clusters (CTP). A: Volcano plot showing fold-change for each gene and their significance level. Genes with an adjusted p-value lower than 0.01 and an absolute $\log_2$ fold change above 2 are colored in orange. Differentially expressed genes included in interferon-dependent pathway are labelled. B: Enrichment of Gene Ontology categories related to Interferon signaling in COVID-19 Transcriptomic Profile 2 (CTP2, n=14 compared to CTP1, n=42). C-E: Networks combining pathways and genes with differential expression between clusters; involving Interferon-dependent lymphoid activation (C) upregulated in CTP1, and B-cell receptor signaling (D) and regulatory T-cell differentiation (E), upregulated in CTP2.

Figure 3. Correlation between genes included in Interferon-dependent pathways. Correlograms (bottom) and gene networks (top) showing correlations with opposite sign between genes in each COVID-19 transcriptomic cluster (CTP, n=42 and 14 for CTP1 and 2 respectively). Only Pearson correlation coefficients with a P-value lower than 0.05 are shown.

Figure 4. Estimated circulating cell populations. A-D: Proportions of blood cells were estimated from RNA-seq using a deconvolution algorithm. E-H: Lymphocyte

subpopulations expressed as percentage of the absolute number of lymphocytes. Points represent individual patient data. In boxplots, bold line represents the median, lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and upper and lower whiskers extend from the hinge to the largest or smallest value no further than 1.5 times the interquartile range. P-values were calculated using a two-tailed Wilcoxon test.

Figure 5. Regulation of gene expression by micro-RNAs. A: Micro-RNAs potentially regulating genes with increased differential expression were identified and a network built. B-H: Counts of hub micro-RNAs (defined as those regulating 3 or more differentially expressed genes) in serum. Points represent individual patient data. In boxplots, bold line represents the median, lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and upper and lower whiskers extend from the hinge to the largest or smallest value no further than 1.5 times the interquartile range. P-values were calculated using a two-tailed Wilcoxon test.

Figure 6. Cluster-specific effects of steroids. Differences in peripheral blood gene expression after 4 days in the ICU between patients treated or not with dexamethasone, stratified by cluster. A-B: Euler diagrams showing the number of genes up- (A) and down-regulated in patients receiving steroids. C: Pathways with divergent activation/suppression response to steroids between clusters.

Figure 7. Intensive Care Unit (ICU) stay. Cumulative incidence of the main outcome (ICU discharge alive and spontaneously breathing), modelled using a competing risk model (with death as a competitive risk) and adjusted by age, sex and need for mechanical ventilation during the ICU stay. The hazard ratio (HR) for COVID-19 transcriptomic profile 2 (CTP-2) with its 95% confidence interval is shown.

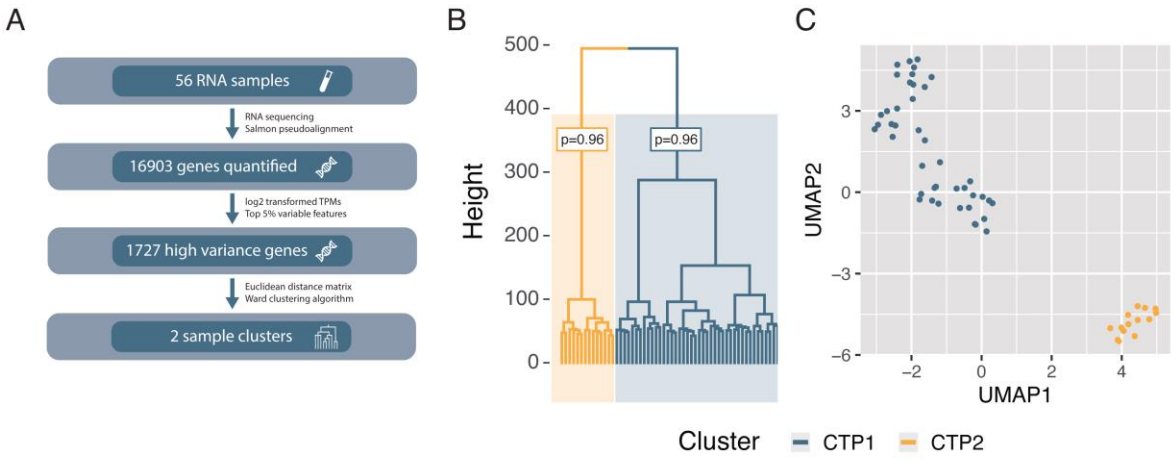
Table 1. Clinical differences between COVID-19 transcriptomic profiles (CTP). BMI: Body mass index. COPD: Chronic Obstructive Pulmonary Disease. APACHE-II: Acute Physiology and Chronic Health disease Classification System II. PBW: Predicted body weight (according to height). PEEP: Positive end-expiratory pressure. IL-6: Interleukin-6. Data are expressed as median (interquartile range) or count (percentage). P-values were calculated using a Wilcoxon test (quantitative data) or Chi-square test (proportions).

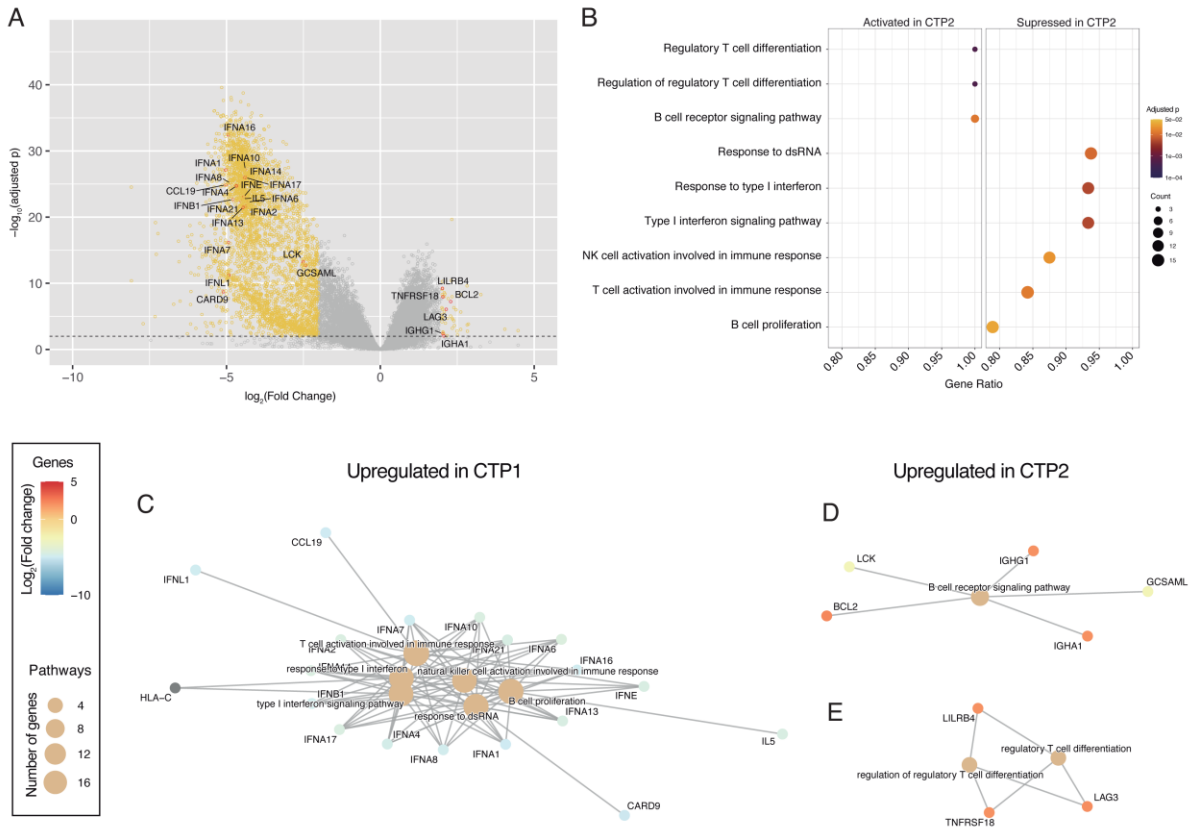
	CTP1 (n=42)	CTP2 (n=14)	p_value
Sex			0.174
Male	36 (86%)	9 (64%)	
Female	6 (14%)	5 (36%)	
Age (years)	69 (63 - 75)	63.5 (59 - 69)	0.147
BMI (Kg/m^2)	29 (25 - 33)	29 (27 - 31)	0.781
Race			0.582
Caucasian	38 (90%)	14 (100%)	
Black	2 (5%)	0	
Latino	2 (5%)	0	
Chronic kidney disease	4 (10%)	0	0.549
COPD	5 (12%)	1 (7%)	1
Liver cirrhosis	1 (2%)	0	1
Arterial hypertension	26 (62%)	6 (43%)	0.35
Diabetes	9 (21%)	3 (21%)	1
Dyslipemia	18 (43%)	6 (43%)	1
Day 1			
APACHE-II score	18 (14 - 21)	16 (13 - 17)	0.120
FiO_2	0.5 (0.4 - 0.6)	0.45 (0.3 - 0.5)	0.438
PaO_2/FiO_2	197 (157 - 245)	188 (151 - 278)	0.863
$PaCO_2$ (mmHg)	43 (39 - 47)	41 (39 - 42)	0.091
Respiratory rate (/min)	18 (16 - 21)	18 (17 - 22)	0.859
pH	7.37 (7.32 - 7.41)	7.42 (7.36 - 7.43)	0.099
Lactate (mEq/L)	1.3 (1.08 - 1.8)	1.1 (0.9 - 1.2)	0.040
Tidal volume (ml)	479 (455 - 504)	500 (475 - 514)	0.499
Tidal volume / PBW (ml/Kg)	7.5 (6.9 - 8.3)	8 (7.5 - 8.7)	0.239
Plateau pressure (cmH_2O)	27 (24 - 29.75)	25 (22 - 29)	0.776
PEEP (cmH_2O)	14 (12 - 15)	12 (10 - 12)	0.088
Driving pressure (cmH_2O)	14 (11 - 15)	15 (12 - 15)	0.568
Compliance (ml/cmH_2O)	36 (31 - 43)	31 (29 - 44)	0.697
Creatinin (mg/dl)	0.92 (0.68 - 1.23)	0.71 (0.59 - 0.97)	0.130
Creatine kinase	96 (60 – 279)	87 (62 – 177)	0.446
Lactate dehydrogenase	440 (396 – 521)	459 (390 – 493)	0.773
Aspartate aminotransferase	47 (37 – 73)	45 (31 – 55)	0.458
Alanine aminotransferase	35 (21 – 55)	29 (20 – 58)	0.893
Procalcitonin (ng/ml)	0.23 (0.14 – 0.6)	0.14 (0.13 – 0.27)	0.250
C Reactive Protein	19 (9 – 24)	16 (2 – 25)	0.699
IL-6 (pg/ml)	113 (54 - 276)	164 (36 – 250)	0.784
Ferritin (ng/ml)	1329 (968 - 1606)	1673 (856 - 2182)	0.576
D-dimer (ng/ml)	1495 (842 - 3304)	1084 (750 - 2126)	0.501
Leukocytes (/μl)	9010 (6750 - 11825)	5440 (4418 - 6453)	0.002

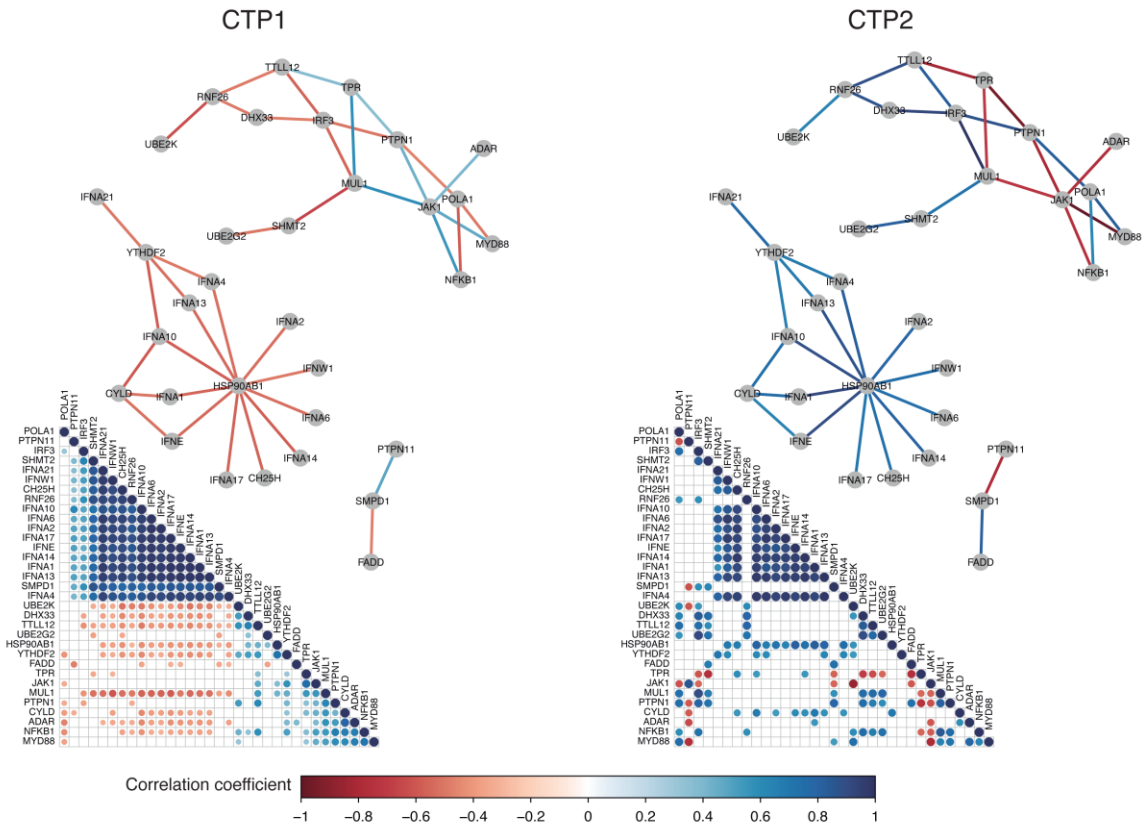
Neutrophils (/μl)	7170 (4990 – 10190)	4180 (3340 – 6200)	0.007
Monocytes (/μl)	330 (180 – 470)	260 (160 – 480)	0.656
Lymphocytes (/μl)	645 (482.5 - 948)	730 (580 - 908)	0.705
Neutrophil-to-Lymphocyte ratio	10.5 (7.8 – 17.3)	7.1 (3.5 – 10.7)	0.010
Days from hospital to ICU admission	2 (0 – 3)	2 (1 – 4)	0.5
Treatments during ICU stay			
Mechanical ventilation	38 (90%)	11 (79%)	0.484
Prone ventilation	23 (61%)	8 (73%)	0.981
Neuromuscular blockade	23 (61%)	6 (55%)	0.643
Extracorporeal membrane oxygenation	1 (3%)	0	1
Vasoactive drugs None One Two or more	 17 (40%) 25 (60%) 0	 6 (43%) 7 (50%) 1 (7%)	0.204
Steroid therapy	19 (45%)	5 (36%)	0.755
ICU evolution			
IL-6 at day 7	54 (11 - 171)	42 (16 - 130)	0.713
Ferritin at day 7	1100 (698 - 1504)	1544 (805 - 1908)	0.745
D-dimer at day 7	2068 (1249 - 4586)	1541 (988 - 3370)	0.422
Ventilator-free days at day 28	12 (0 - 19)	19 (9 – 23)	0.050

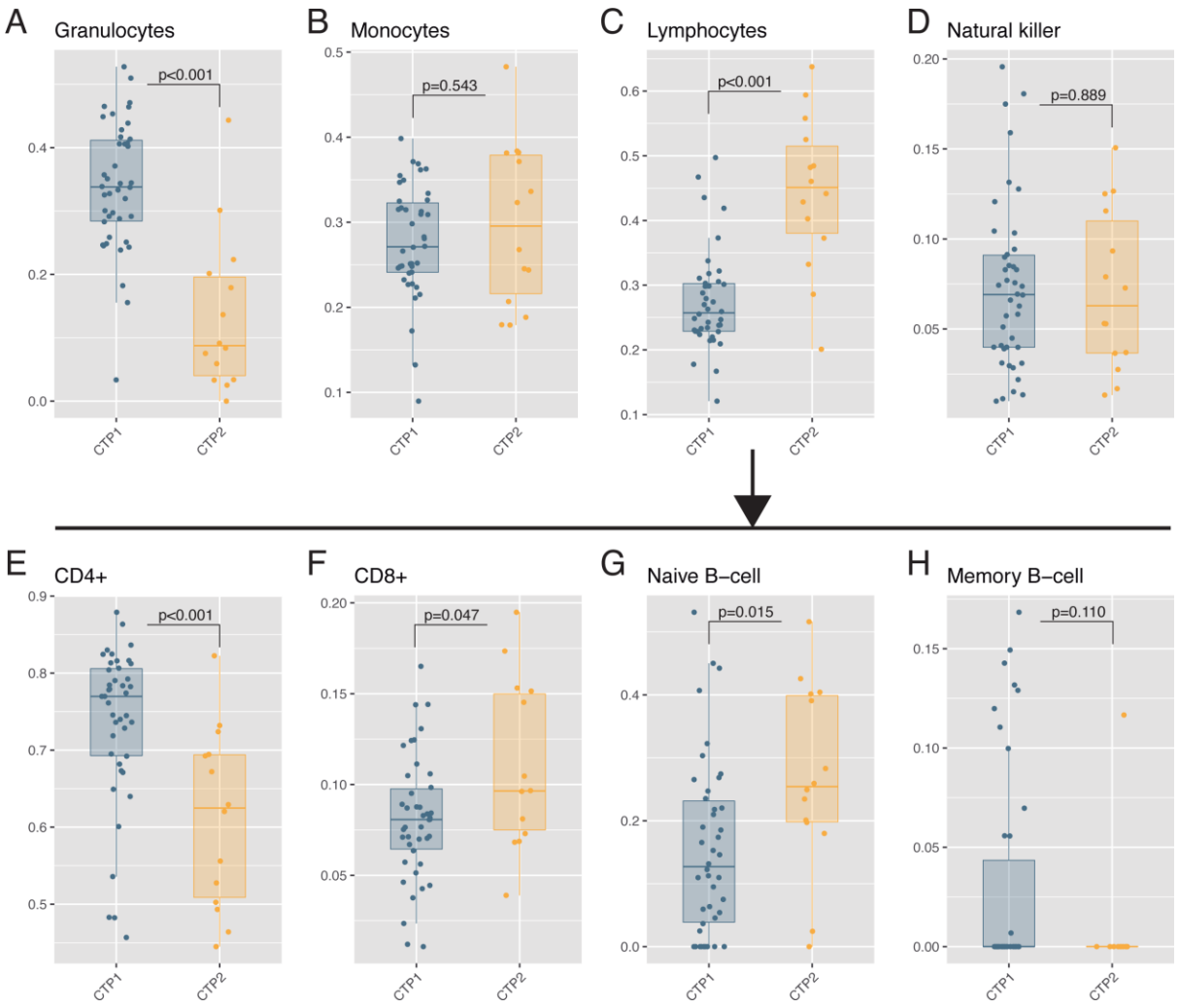
Table 2. Clinical data and outcomes in the validation cohort. APACHE-II: Acute Physiology and Chronic Health disease Classification System II. SOFA: Sequential Organ Failure Assessment. VFD: Ventilator-free days. P-values were calculated using a Wilcoxon test (quantitative data) or Chi-square test (proportions).

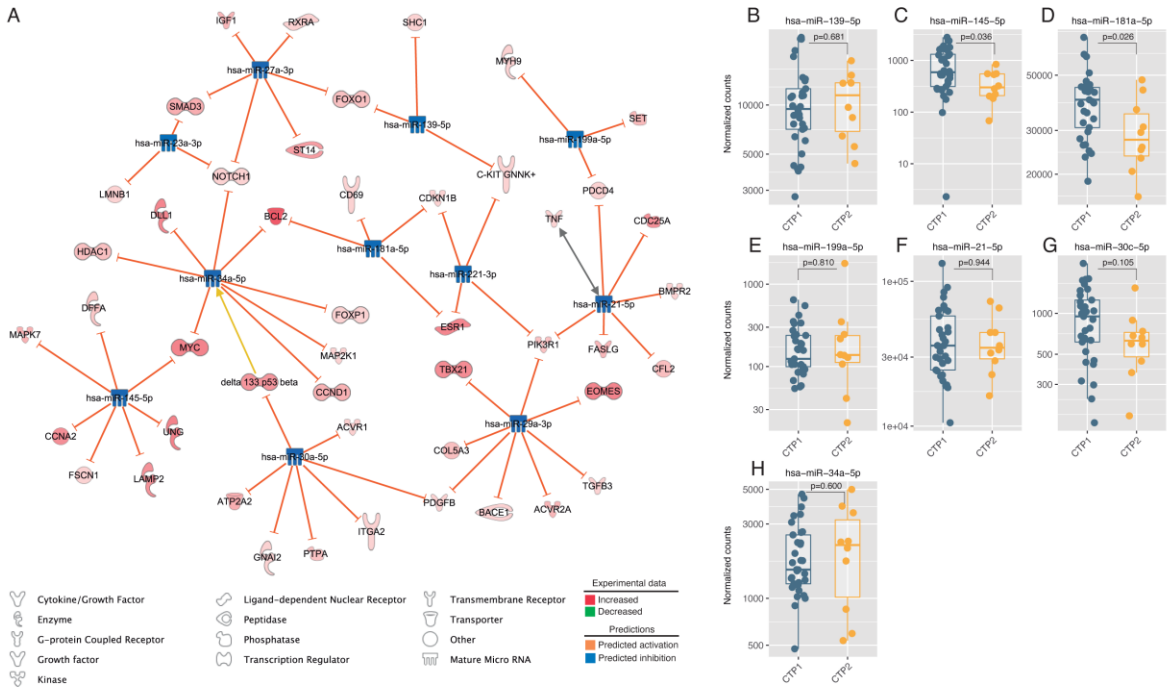
	CTP1	CTP2	p_value
Validation cohort 1			
Sample size	13	37	
Transcriptomic score	216 (197 - 228)	365 (311 - 470)	
Age (years)	63 (55 - 73)	64 (55 - 72)	0.842
Sex (male/female)	8 / 5	25 / 12	0.741
APACHE-II score	23 (20 - 34)	21 (14 - 25)	0.097
SOFA score	7 (6 - 13)	8 (6 - 10)	0.35
VFDs at day 28	0 (0 - 20)	18 (2 - 28)	0.016
Zero VFDs at day 28	8 (62%)	8 (22%)	0.014
Validation cohort 2			
Sample size	22	38	
Transcriptomic score	1430 (1215 – 1506)	2194 (1782 – 2503)	
Age ≥ 50 y	18 (	31	1
Sex (male/female)	13 / 9	21 / 17	0.986
Death at day 28	7	3	0.042

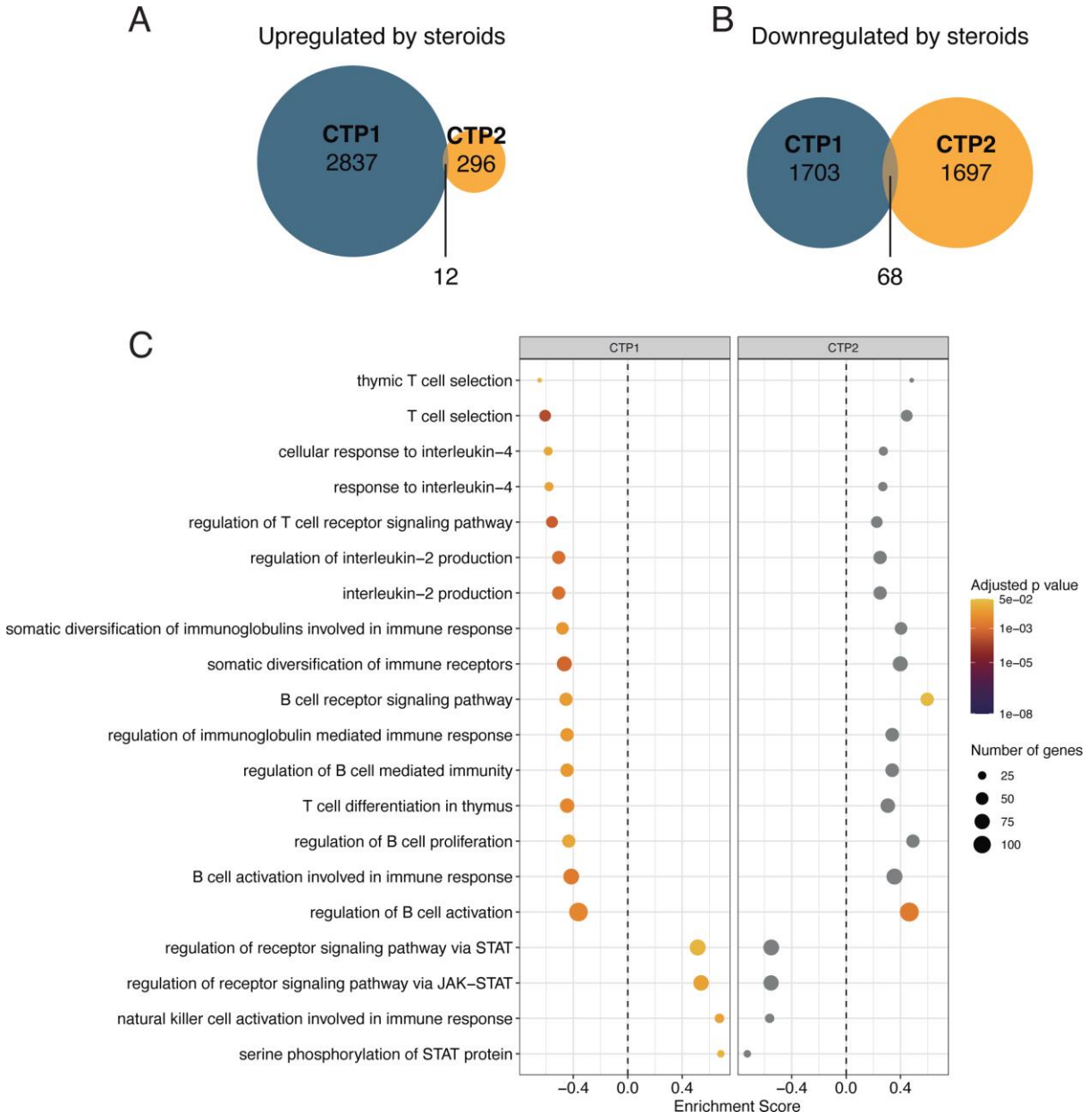

A
56 RNA samples
RNA sequencing
Salmon pseudoalignment
16903 genes quantified
log2 transformed TPMs
Top 5% variable features
1727 high variance genes
Euclidean distance matrix
Ward clustering algorithm
2 sample clusters
B
Height
500
400
300
200
100
0
p=0.96
p=0.96
C
UMAP2
3
0
−3
−6
−2
0
2
4
UMAP1
Cluster
CTP1
CTP2

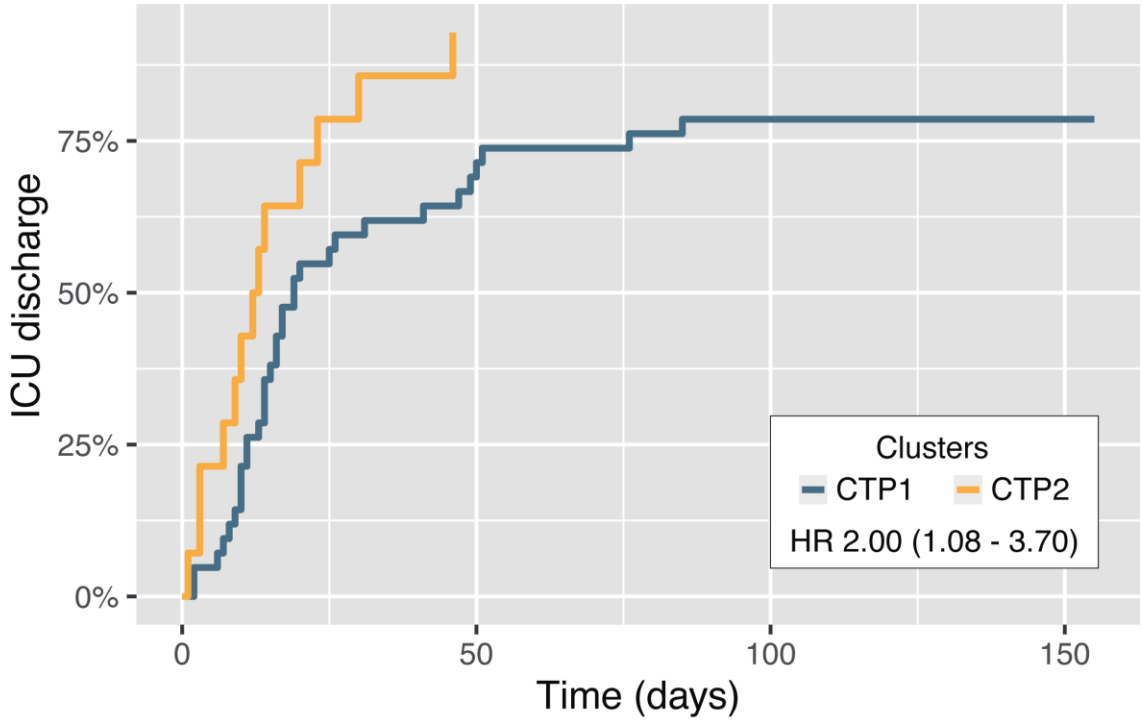

A
B
Activated in CTP2
Supressed in CTP2
Regulatory T cell differentiation
Regulation of regulatory T cell differentiation
B cell receptor signaling pathway
Response to dsRNA
Response to type I interferon
Type I interferon signaling pathway
NK cell activation involved in immune response
T cell activation involved in immune response
B cell proliferation
Gene Ratio
Adjusted p
Count
IFNA16
IFNA1
IFNA10
IFNA14
IFNA8
IFNE
IFNA17
CCL19
IFNA4
IL5
IFNA6
IFNB1
IFNA21
IFNA2
IFNA13
IFNA7
LCK
GCSAML
IFNL1
CARD9
LILRB4
TNFRSF18
BCL2
LAG3
IGHG1
IGHA1
log2(Fold Change)
−log10(adjusted p)
Genes
Log2(Fold change)
Pathways
Number of genes
C
Upregulated in CTP1
CCL19
IFNL1
IFNA7
IFNA10
T cell activation involved in immune response
IFNA2
IFNA21
IFNA5
IFNA16
IFNA1
response to type I interferon
natural killer cell activation involved in immune response
HLA−C
IFNB1
type I interferon signaling pathway
IFNE
B cell proliferation
IFNA13
response to dsRNA
IFNA17
IFNA4
IL5
IFNA8
IFNA1
CARD9
D
Upregulated in CTP2
LCK
IGHG1
B cell receptor signaling pathway
GCSAML
BCL2
IGHA1
E
LILRB4
regulatory T cell differentiation
regulation of regulatory T cell differentiation
LAG3
TNFRSF18


CTP1
CTP2
TTLL12
TPR
RNF26
DHX33
IRF3
UBE2K
PTPN1
ADAR
MUL1
IFNA21
POLA1
JAK1
UBE2G2
SHMT2
MYD88
YTHDF2
NFKB1
IFNA4
IFNA13
IFNA2
IFNA10
IFNW1
CYLD
IFNA1
HSP90AB1
IFNA6
IFNE
IFNA14
IFNA17
CH25H
PTPN11
SMPD1
FADD
Correlation coefficient
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1


A
Granulocytes
p<0.001
B
Monocytes
p=0.543
C
Lymphocytes
p<0.001
D
Natural killer
p=0.889
E
CD4+
p<0.001
F
CD8+
p=0.047
G
Naive B−cell
p=0.015
H
Memory B−cell
p=0.110
CTP1
CTP2


A
IGF1
RXRA
SHC1
MYH9
hsa-miR-27a-3p
FOXO1
hsa-miR-139-5p
SET
hsa-miR-199a-5p
SMAD3
hsa-miR-23a-3p
ST14
NOTCH1
C-KIT GNNK+
PDCD4
LMNB1
DLL1
CD69
CDKN1B
BCL2
hsa-miR-181a-5p
TNF
CDC25A
HDAC1
hsa-miR-34a-5p
hsa-miR-221-3p
BMPR2
hsa-miR-21-5p
DFFA
FOXP1
ESR1
PIK3R1
FASLG
MAPK7
MYC
MAP2K1
CFL2
TBX21
delta 133 p53 beta
CCND1
hsa-miR-145-5p
EOMES
hsa-miR-29a-3p
CCNA2
UNC
ACVR1
COL5A3
hsa-miR-30a-5p
FSCN1
LAMP2
TGFB3
ATP2A2
PDGFB
BACE1
ACVR2A
ITGA2
GNAI2
PTPA
Cytokine/Growth Factor
Enzyme
G-protein Coupled Receptor
Growth factor
Kinase
Ligand-dependent Nuclear Receptor
Peptidase
Phosphatase
Transcription Regulator
Transmembrane Receptor
Transporter
Other
Mature Micro RNA
Experimental data
Increased
Decreased
Predictions
Predicted activation
Predicted inhibition
B
hsa-miR-139-5p
p=0.681
C
hsa-miR-145-5p
p=0.036
D
hsa-miR-181a-5p
p=0.026
E
hsa-miR-199a-5p
p=0.810
F
hsa-miR-21-5p
p=0.944
G
hsa-miR-30c-5p
p=0.105
H
hsa-miR-34a-5p
p=0.600
Normalized counts
CTP1
CTP2


A
Upregulated by steroids
CTP1
2837
CTP2
296
12
B
Downregulated by steroids
CTP1
1703
CTP2
1697
68
C
CTP1
CTP2
thymic T cell selection
T cell selection
cellular response to interleukin−4
response to interleukin−4
regulation of T cell receptor signaling pathway
regulation of interleukin−2 production
interleukin−2 production
somatic diversification of immunoglobulins involved in immune response
somatic diversification of immune receptors
B cell receptor signaling pathway
regulation of immunoglobulin mediated immune response
regulation of B cell mediated immunity
T cell differentiation in thymus
regulation of B cell proliferation
B cell activation involved in immune response
regulation of B cell activation
regulation of receptor signaling pathway via STAT
regulation of receptor signaling pathway via JAK−STAT
natural killer cell activation involved in immune response
serine phosphorylation of STAT protein
−0.4
0.0
0.4
Enrichment Score
Adjusted p value
5e−02
1e−03
1e−05
1e−08
Number of genes
25
50
75
100



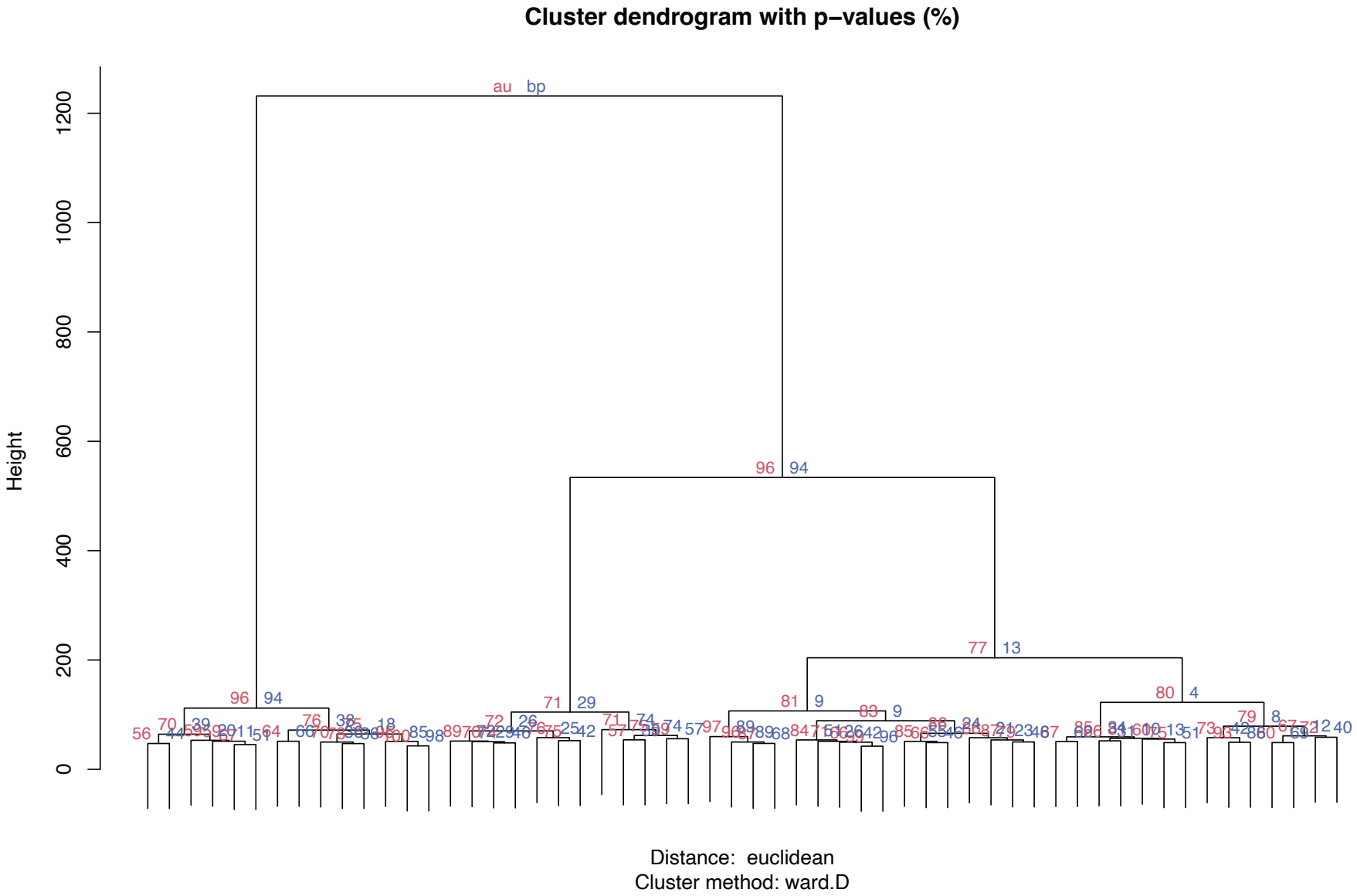
Time ▶		0	25	50	75	100	125	150
N. at risk	CTP1	42	15	6	4	1	1	1
	CTP2	14	3	0	0	0	0	0

Transcriptomic clustering of critically ill COVID-19 patients

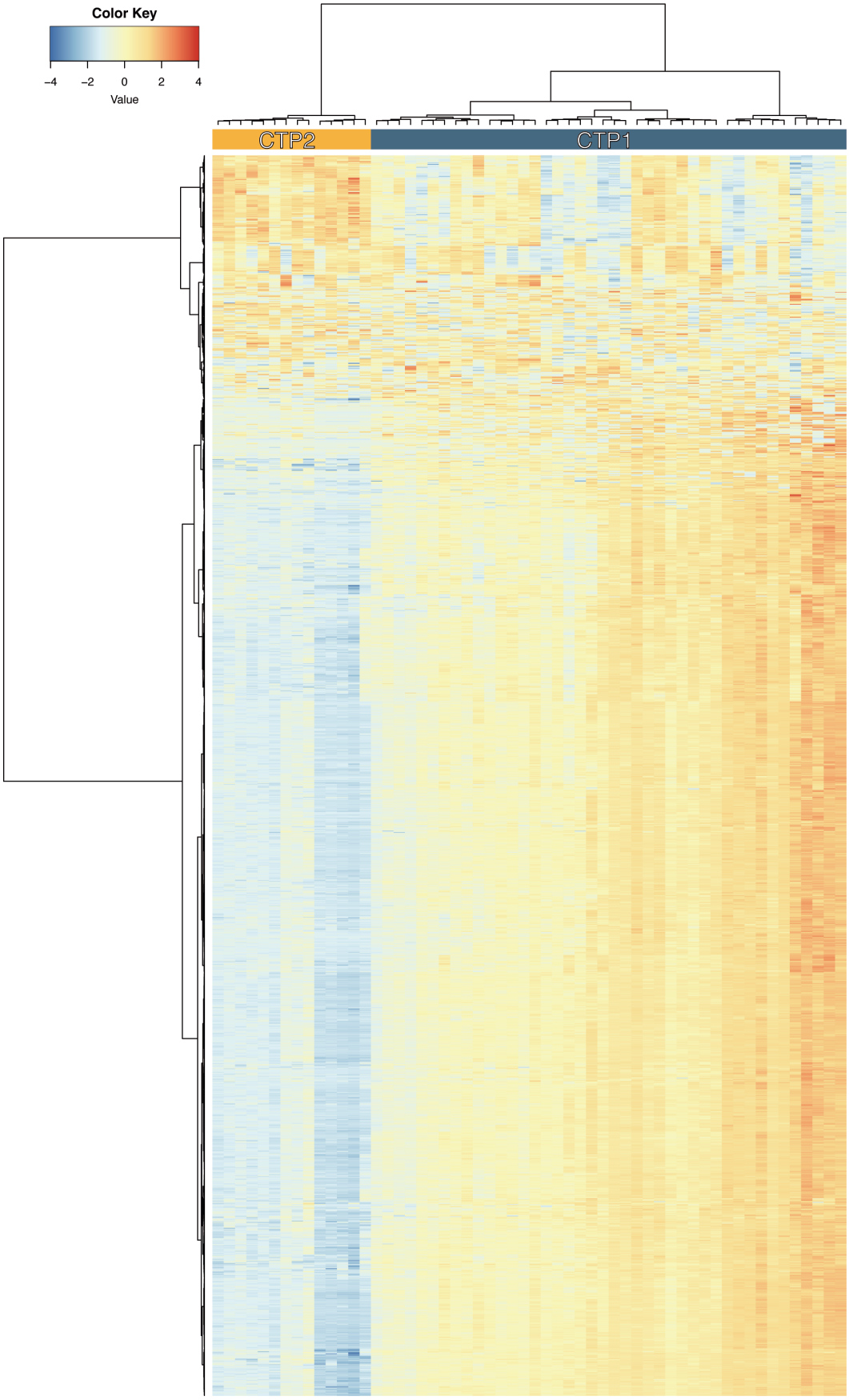
Cecilia López-Martínez, Paula Martín-Vicente, Juan Gómez de Oña, Inés López-Alonso, Helena Gil-Peña, Elías Cuesta-Llavona, Margarita Fernández-Rodríguez, Irene Crespo, Estefanía Salgado del Riego, Raquel Rodríguez-García, Diego Parra, Javier Rodríguez-Carrio, Alberto Dávalos, Luis A Chapado, Eliecer Coto, Guillermo M Albaiceta, Laura Amado-Rodríguez

Online data supplement

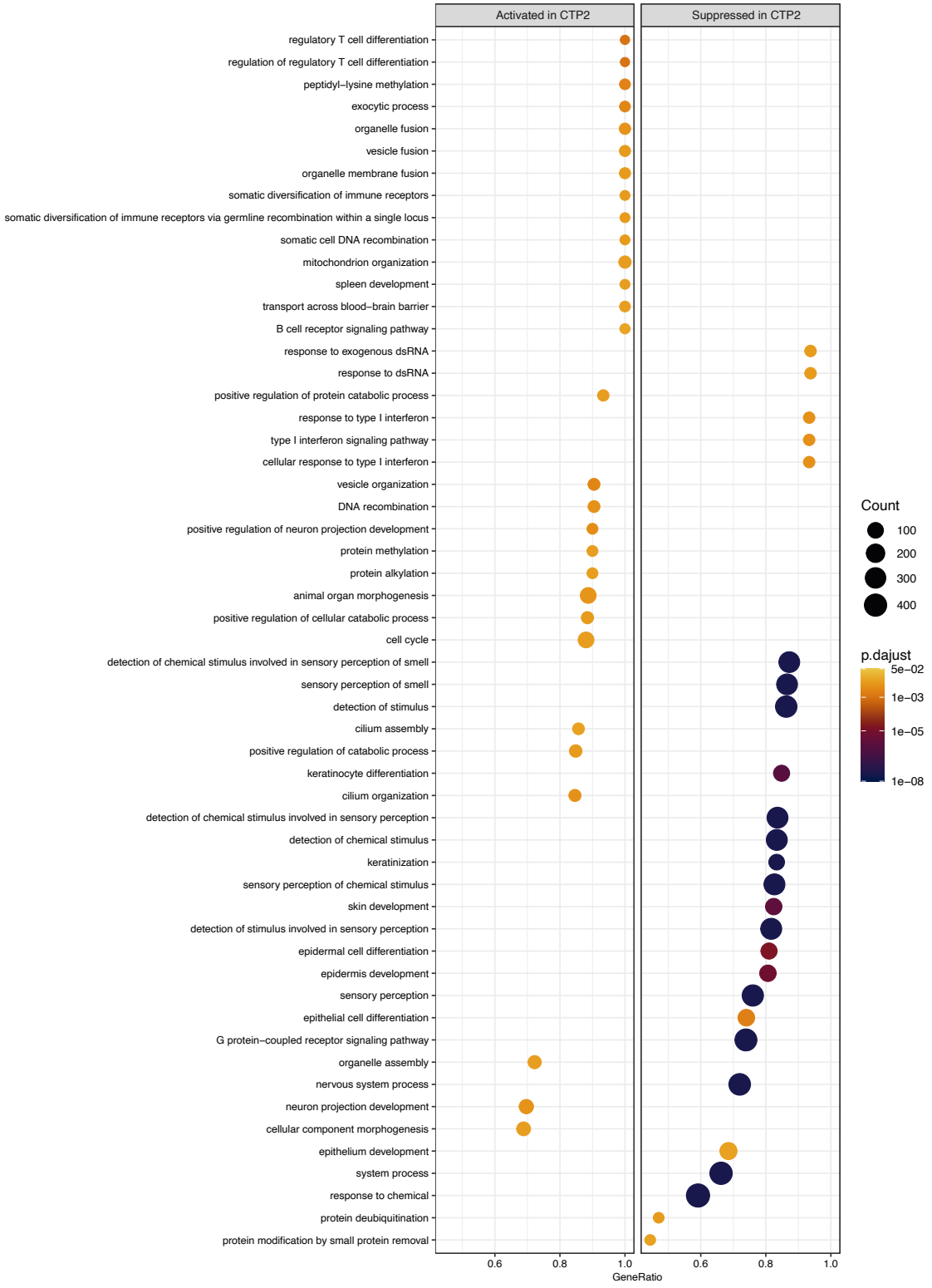
Supplementary figure 1. Hierarchical clustering tree showing the p values for each cluster, calculated using the pvclust package for R. AU (red): Approximately unbiased p-values. BP (blue): Bootstrap probability.



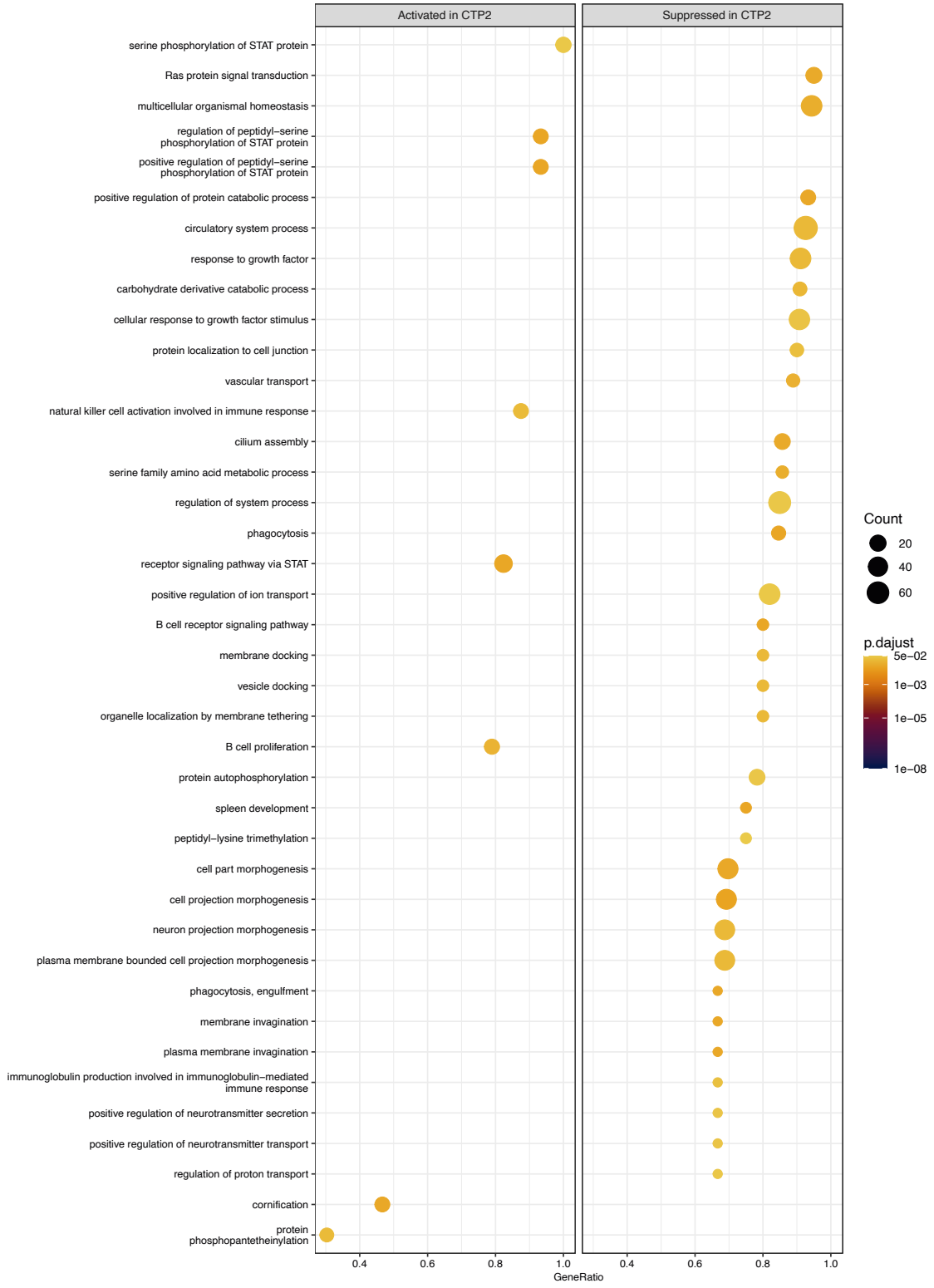
Supplementary figure 2. Heatmap showing differences in expression between COVID transcriptomic profiles (CTP) in the 1727 genes used for clustering.



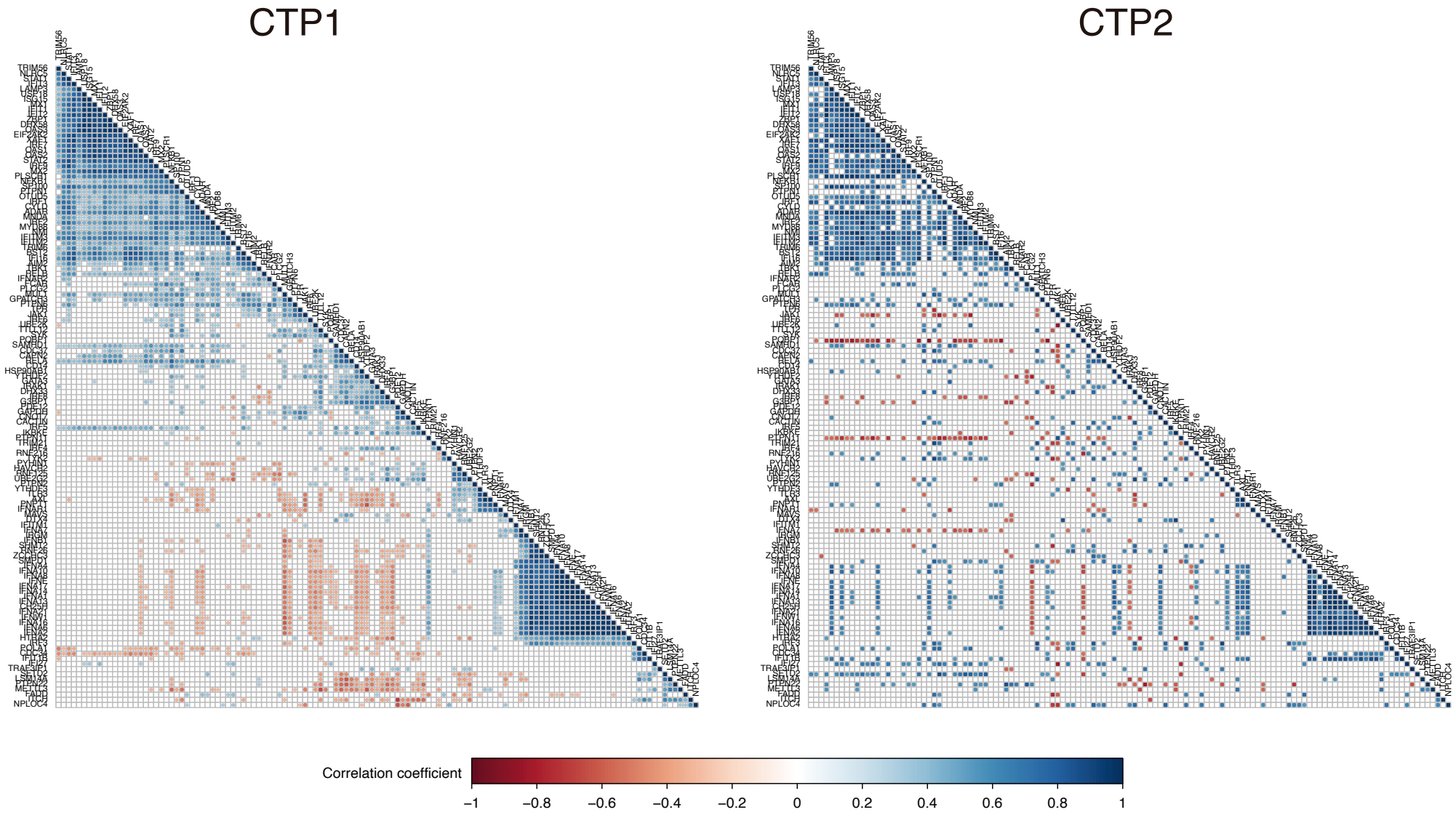
Supplementary figure 3. Biological processes with significant differences in gene enrichment between COVID transcriptomic profiles (CTP).



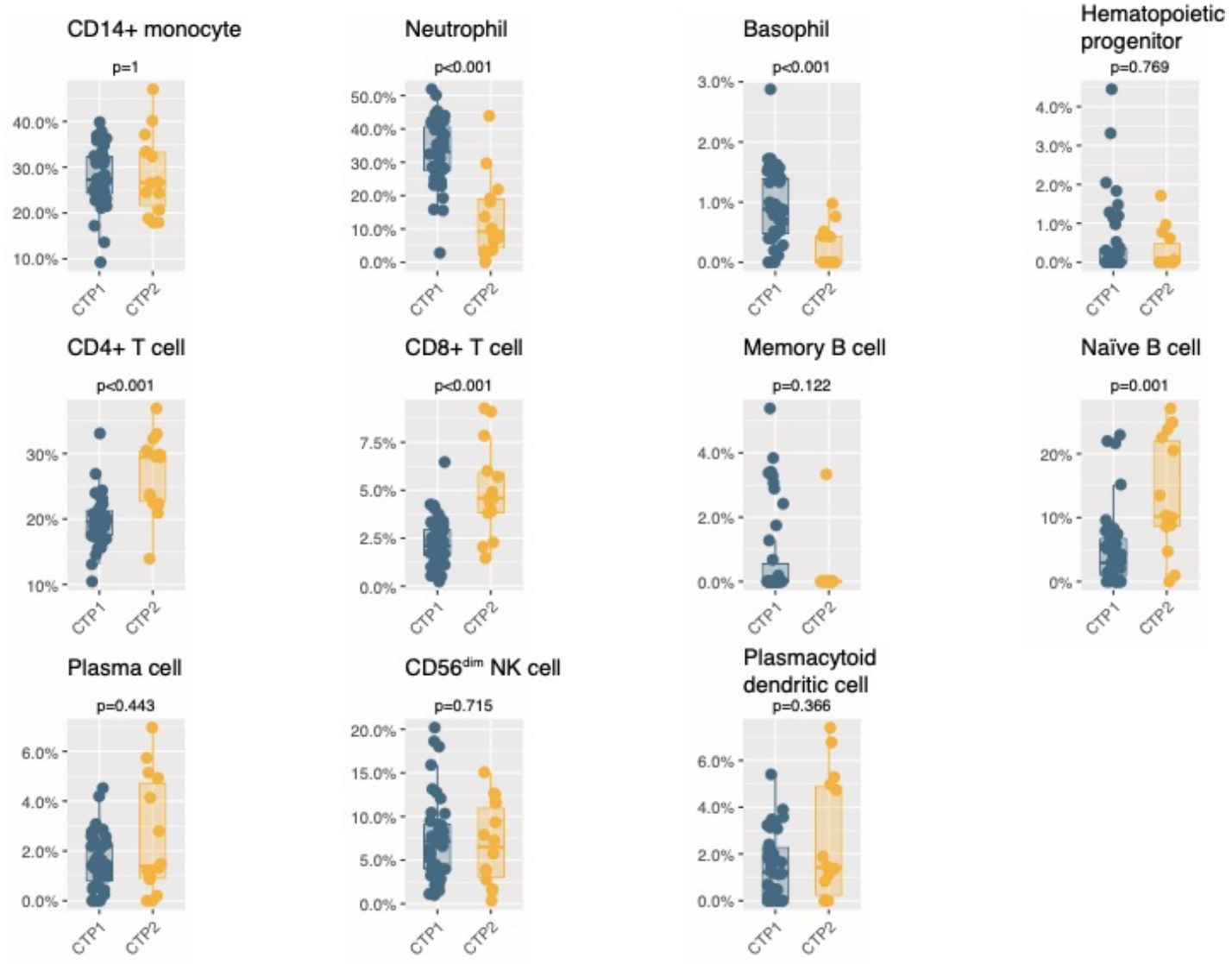
Supplementary figure 3 (cont). Biological processes with significant differences in gene enrichment between COVID transcriptomic profiles (CTP).



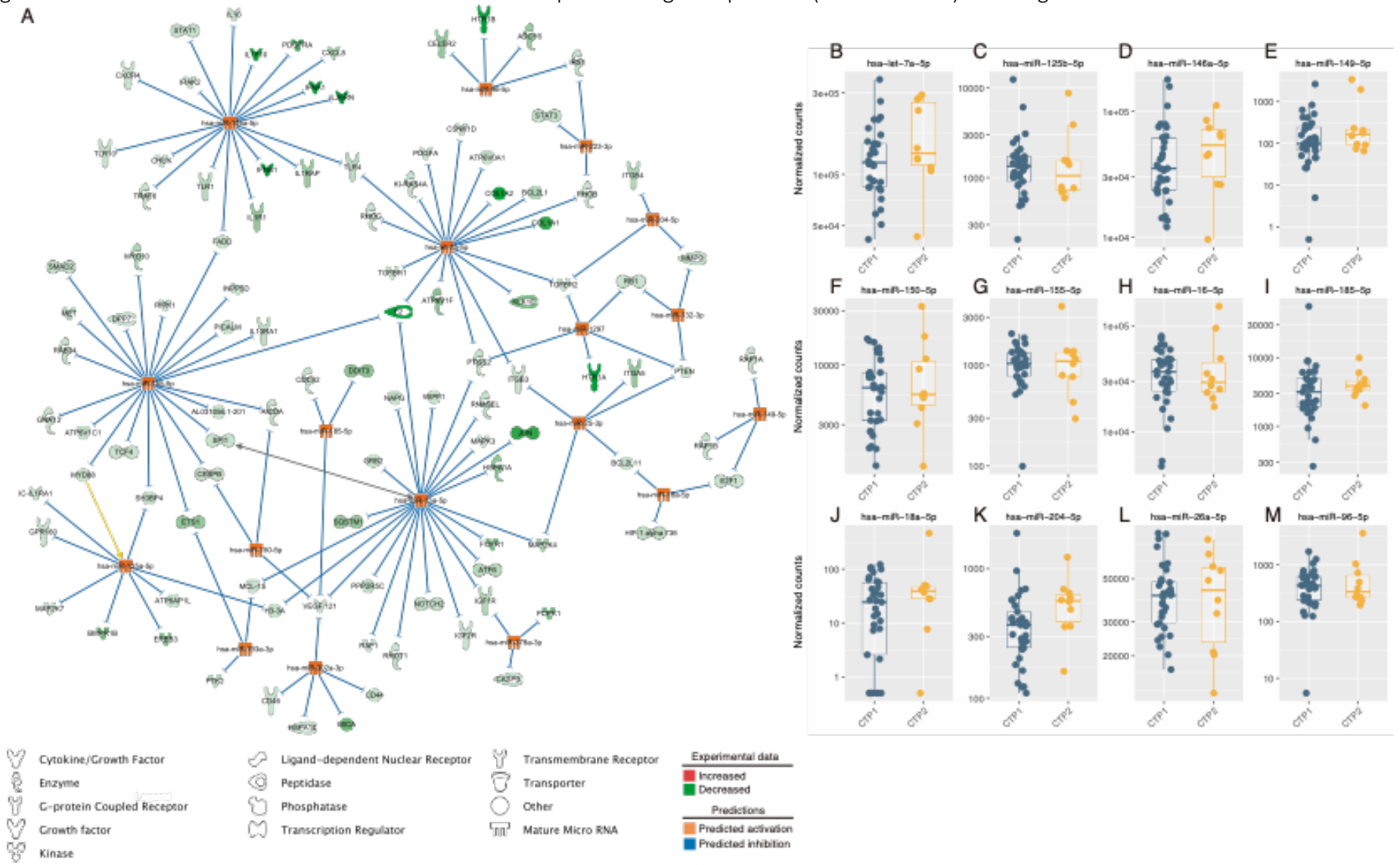
Supplementary figure 4. Correlations between genes in each COVID transcriptomic profile (CTP).



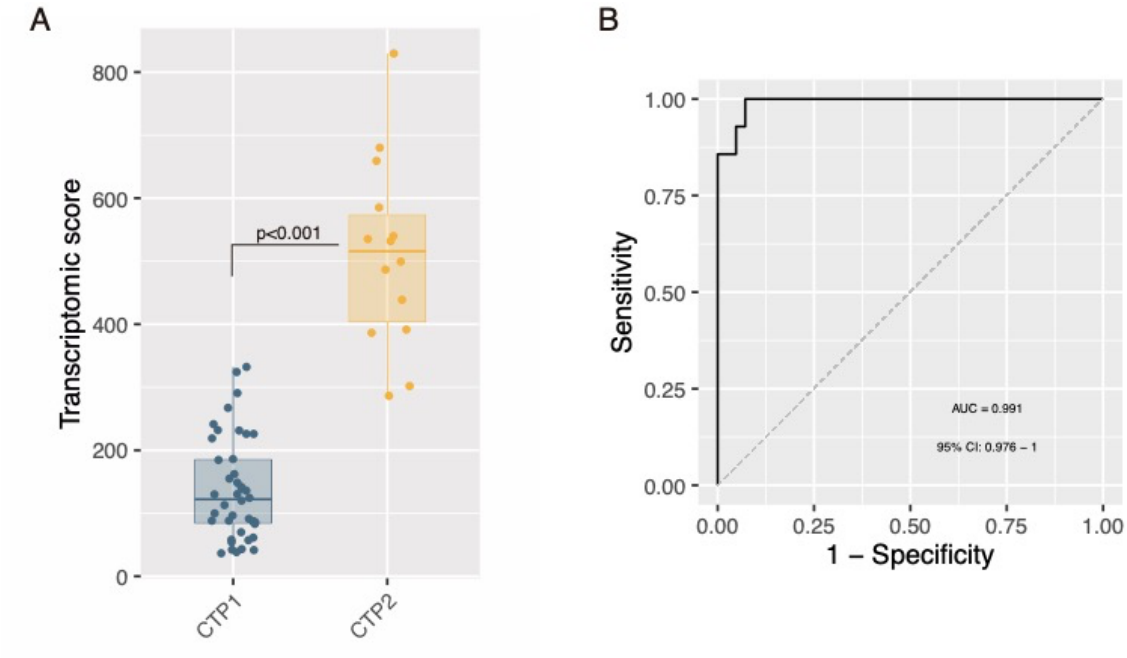
Supplementary figure 5. Differences in estimated cell populations between clusters. CTP: COVID transcriptomic profile. Points represent individual patient data. In boxplots, bold line represents the median, lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and upper and lower whiskers extend from the hinge to the largest or smallest value no further than 1.5 times the interquartile range. P-values were calculated using a two-tailed Wilcoxon test.



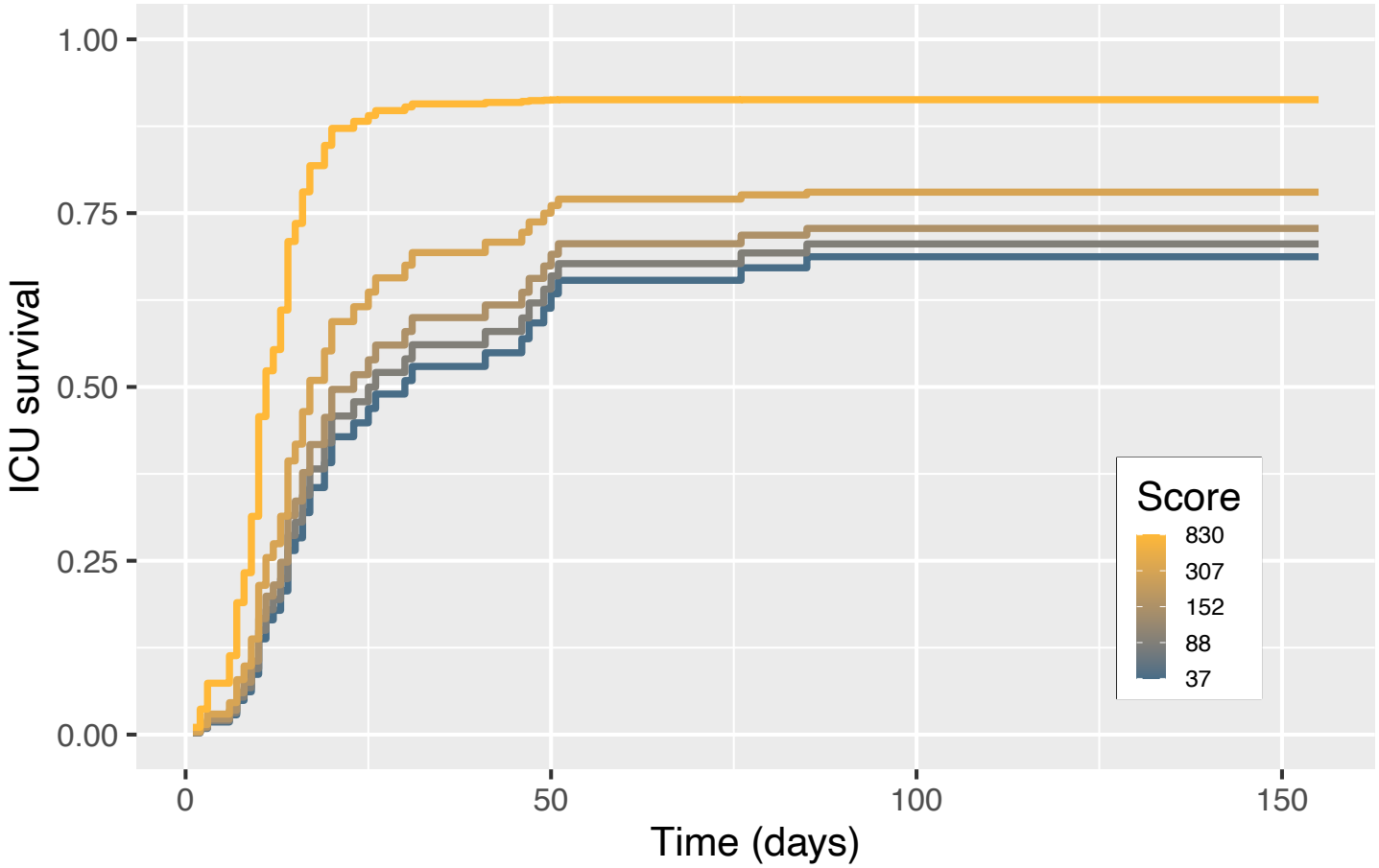
Supplementary figure 6. A: MicroRNA/gene network showing predicted upregulated miRNAs related to downregulated genes in patients with COVID transcriptomic profile (CTP) 2. B-M: Normalized counts of each miRNA. Points represent individual patient data. In boxplots, bold line represents the median, lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and upper and lower whiskers extend from the hinge to the largest or smallest value no further than 1.5 times the interquartile range. All p values (Wilcoxon test) were higher than 0.05.



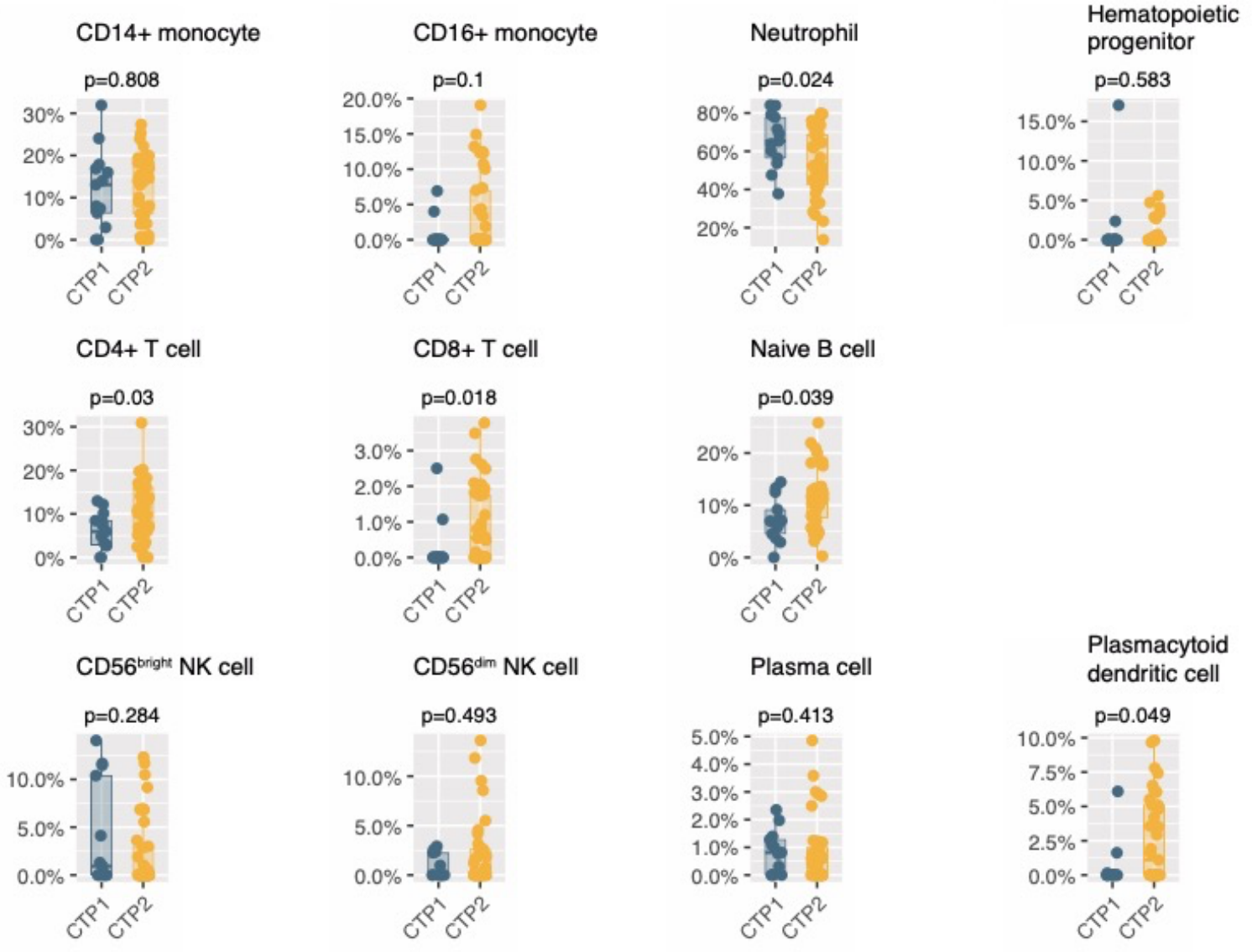
Supplementary figure 7. A: Transcriptomic scores calculated in the training sets. B: ROC curve corresponding to the diagnostic accuracy of the transcriptomic score to identify COVID transcriptomic profiles.



Supplementary figure 8. Cumulative incidence curves of ICU discharge alive and spontaneously breathing, modelled using the developed Cox model, for increasing transcriptomic scores (using percentiles 0, 25, 50, 75 and 100 of the sample), keeping constant all the other factors included in the model (age, sex, need for intubation).



Supplementary figure 9. Estimated cell populations according to COVID transcriptomic profiles (CTP) in the first validation cohort. Points represent individual patient data. In boxplots, bold line represents the median, lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and upper and lower whiskers extend from the hinge to the largest or smallest value no further than 1.5 times the interquartile range. P-values were calculated using a two-tailed Wilcoxon test.



Supplementary figure 10. Estimated cell populations according to COVID transcriptomic profiles (CTP) in the second validation cohort. Points represent individual patient data. In boxplots, bold line represents the median, lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and upper and lower whiskers extend from the hinge to the largest or smallest value no further than 1.5 times the interquartile range. P-values were calculated using a two-tailed Wilcoxon test.