Host transcriptomic profiling of COVID-19 patients with mild, moderate, and severe clinical outcomes

Ruchi Jain[1], Sathishkumar Ramaswamy[1], Divinlal Harilal[1], Mohammed Uddin[2,3], Tom Loney[2], Norbert Nowotny[2,4], Hanan Alsuwaidi[2], Rupa Varghese[5], Zulfa Deesi[5], Abdulmajeed Alkhajeh[6], Hamda Khansaheb[6], Alawi Alsheikh-Ali[2], Ahmad Abou Tayoun[1,2*]

[1]Al Jalila Genomics Center, Al Jalila Children's Hospital, Dubai, United Arab Emirates.

[2]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates.

[3]The Centre for Applied Genomics (TCAG), The Hospital for Sick Children, Toronto, ON, Canada.

[4]Institute of Virology, University of Veterinary Medicine Vienna, Vienna, Austria.

[5]Microbiology and Infection Control Unit, Pathology and Genetics Department, Latifa Women and Children Hospital, Dubai Health Authority, Dubai, United Arab Emirates.

[6]Medical Education & Research Department, Dubai Health Authority, Dubai, United Arab Emirates.

*Corresponding Author: Ahmad Abou Tayoun (Ahmad.Tayoun@ajch.ae)

Financial Support and Competing Interests Statement

Abstract

Characterizing key molecular and cellular pathways involved in COVID-19 is essential for disease prognosis and management. We perform shotgun transcriptome sequencing of human RNA obtained from nasopharyngeal swabs of patients with COVID-19, and identify a molecular signature associated with disease severity. Specifically, we identify globally dysregulated immune related pathways, such as cytokine-cytokine receptor signaling, complement and coagulation cascades, JAK-STAT, and TGF- β signaling pathways in all, though to a higher extent in patients with severe symptoms. The excessive release of cytokines and chemokines such as *CCL2*, *CCL22*, *CXCL9* and *CXCL12* and certain interferons and interleukins related genes like *IFIH1*, *IFI44*, *IFIT1* and *IL10* were significantly higher in patients with severe clinical presentation compared to mild and moderate presentations. Moreover, early induction of the TGF- β signaling pathway might be the primary cause of pulmonary fibrosis in patients with severe disease. Differential gene expression analysis identified a small set of regulatory genes that might act as strong predictors of patient outcome. Our data suggest that rapid transcriptome analysis of nasopharyngeal swabs can be a powerful approach to quantify host molecular response and may provide valuable insights into COVID-19 pathophysiology.

Key Words: Host Transcriptome, COVID-19, Clinical Outcomes, Sequenicng

Introduction

Since the first case reported in December 2019 in Wuhan, SARS-CoV-2 has spread very quickly across 188 different countries leading to over 31,887,485 COVID-19 cases and 976,789 associated deaths worldwide as of 24th September 2020.[1] The common signs and symptoms of SARS-CoV-2 include fever, cough, fatigue, shortness of breath, and chest CT abnormalities.[2] Some patients with COVID-19 quickly develop severe pulmonary symptoms including acute respiratory distress syndrome (ARDS), pulmonary edema, intense kidney injury and multiple organ failure[3] whereas other patients present with no symptoms or with only mild disease.[4] Pathogen load and prevalence of infections are higher in males than females demonstrating that sex physiology plays a role in infectious disease pathogenesis.[5] Meng and colleagues[6] showed that men who died of COVID-19 had elevated levels of systemic inflammatory markers such as neutrophil-to-lymphocyte ratio and C-reactive protein. Nevertheless, the mechanism behind increased mortality among older adults and males with COVID-19 remains speculative.

Despite the worldwide spreading, the host immune response against SARS-CoV-2 infection remains poorly characterized. Dysregulation of host immune response and activation of inflammatory cytokines, known as the "cytokine storm", is associated with disease severity and poor prognosis.[7, 8, 9] A recent host transcriptome study on patients with COVID-19 revealed distinct host inflammatory cytokine profiles and highlighted the association between SARS-CoV-2 pathogenesis and excessive release of cytokine such as CCL2/MCP-1, CXCL10/IP-10, CCL3/MIP-1A, and CCL4/MIP1B.[10] Additionally, Lieberman *et al.* showed upregulation of antiviral factors such as OAS1-3 and IFIT1-3, and Th1 chemokines CXCL9/10/11 upon SARS-CoV2 induced antiviral response. This study also showed immune responses may underlie disparities between males and females, and among the elderly compared to younger age groups.[11]

At present, it is not known if the host gene expression profile varies among patients with mild, moderate, or severe clinical outcomes. Identification of such transcriptomic differences can be useful for predicting COVID-19 outcomes and for better management and earlier interventions especially for patient groups with severe outcomes. In this study, we performed RNA sequencing on nasopharyngeal tissue of patients with mild, moderate, and severe disease to characterize transcriptomic regulation and immune response differences among

these patients, and to identify markers for early identification of patients with possible severe disease outcomes.

Methods

Patient cohort and ethics statement

This study was approved by the Dubai Scientific Research Ethics Committee - Dubai Health Authority (approval number #DSREC-04/2020_02). The Ethics committee waived the requirement for informed consent since this study was part of a public health surveillance and outbreak investigation in the UAE. The electronic medical records of patients with laboratory confirmed SARS-CoV-2 were reviewed and important clinical data were extracted using the World Health Organization (WHO) case report form.

Sample preparation, RNA isolation, library construction and sequencing

RNA was extracted from nasopharyngeal swabs using the QIAamp Viral RNA Mini or the 213 EZ1 DSP Virus Kits (Qiagen, Hilden, Germany). All patients in this study tested positive for SARS-CoV-2 by RT-qPCR performed at Dubai Health Authority Hospitals. RNA libraries were then prepared for shotgun transcriptome sequencing using the TruSeq Stranded Total RNA Library kit from Illumina (San Diego, CA, USA) as previously described.[12,13,14] Libraries were then sequenced using the NovaSeq SP Reagent kit (2 X 150 cycles) from Illumina (San Diego, CA, USA) to generate a minimum of 15M reads per sample (**Supplementary Table 1**).

Data analysis

The sequencing read quality was checked using FastQC v0.11.8[15] and MultiQC v1.6.[16] High quality reads (Q ≥ 30) were first mapped to rRNA sequences to remove potential rRNA reads using hisat v2-2.1.0[17] with the default parameter. The remaining reads were then mapped to the GRCh37 (hg19). The unmapped reads were then mapped to the SARS-CoV-2 genome (GenBank Accession number: NC_045512.2) using BWA v0.7.17. PCR duplicates were removed with Picard MarkDuplicates program v2.18.17.[18] Samples with more than 6M reads aligned to the GRCh37 (hg19) were considered for further analysis. RNA sequencing data obtained from nasopharyngeal swabs of samples confirmed to be COVID-19 negative by RT-qPCR, hereafter referred to as controls (n = 32), were downloaded from GSE152075.[11]

The number of reads mapped to each gene in the genome (GRCh37) was calculated using the FeatureCounts program in the SubReads package v2.0.1.[19] DESeq2 package v1.28.1[20] was applied to perform batch effects and normalization. In brief, DESeq2 uses the median of ratios method to normalize data and estimate size factors (which control for differences in the library size of the sequencing experiment). Further, gene expression change was calculated with respect to control for each groups using Wald test and p-values and $\log_2$ fold change was extracted. The resulting genes p-value was adjusted using the Benjamini and Hochberg method. These adjusted p-value (adj p-value) are calculated as: gene p-value $\times$ (m/i)) where m is the total number of genes, i is the gene p-value rank. Here we have considered a fraction of 5% false positives acceptable hence the genes with adj p-value < 0.05 were called as significant.

Pathway enrichment analysis was performed using the clusterProfiler package v 3.16.0[21] to identify shared pathways among DEGs. Pathways with adj p-value < 0.05 were reported as significant. Heatmaps were generated using Morpheus[22] and volcano plots were generated using VolcaNoseR package.[23] Both case and control sequencing data were normalized to counts per million (CPM), which is calculated by normalizing the read counts for a given gene by the total counts per sample. CPM values were used to compare expression between certain genes and violin plots were generated using GraphPad Prism v8.0. [24] Two tailed Mann–Whitney U test P-values are reported. A schematic illustration of data analysis is represented in **Figure 1**.

Results

Patient Cohort

Our patient cohort consisted of fifty patients with COVID-19 (36% female) who tested positive for SARS-CoV-2 by RT-qPCR with a mean age of 40.9 years (ranging from 3 – 70 years) at diagnosis. Cases were categorized into mild, moderate, and severe as previously described.[12] Briefly, mild cases were asymptomatic or had mild non-life-threatening symptoms, while moderate cases presented with symptoms (such as persistent fever) requiring medical attention and/or hospitalization. Severe cases presented with advanced disease and pneumonia requiring admission to intensive care units and life-support treatment (such as mechanical ventilation). In total, 37 patients (21 males and 16 females; mean age

36.4 years) had mild disease, 10 patients (9 males and 1 female; mean age 49.3 years) had moderate presentations, while 3 patients (2 males and 1 female; mean age 68.0 years) had severe/critical disease. All patients were non-smokers. All 3 patients with severe disease had chronic health issues (hypertension, chronic kidney disease and/or diabetes mellitus) (**Table 1**).

Global transcriptomic changes in patients with COVID-19

To understand the host mechanisms of SARS-CoV-2 infections, we explored transcriptomic profiling of patients with COVID-19 using RNA extracted from nasopharyngeal samples (see Methods). We first mapped high quality reads to the human genome and the SARS-CoV-2 genome (GenBank accession number: NC_045512.2) and summarized sequencing statistics in **Supplementary Table 1**. On average, ~2% (ranging from 0.1% - 13.6%) of the reads mapped to the SARS-CoV-2 genome confirming the presence of the virus, albeit coverage over the viral genome varied across patients most likely due to differences in viral loads. On the other hand, ~65% (ranges from 24.3% - 91.0%) of the RNA reads mapped to the human genome, enabling host transcriptomic analysis.

To detect global signature genes in patients with COVID-19 we combined transcriptomic data from patients with mild, moderate, and severe disease and made a comprehensive cohort called "COVID" to compare against controls for DEG analysis (see Methods). We detected 3572 upregulated genes and 445 downregulated genes with adj p-value < 0.05 and fold-change cutoff > 2. Several interferon, cytokine and immune-related genes, such as *CXCL5*, *CXCL12*, *CCL2*, *IFIH1*, *IFI44*, *IFIT1* and *IL10* were upregulated, whereas metabolic pathways or housekeeping genes, like *RPL41*, *RPL17*, *SLC25A6*, *CALM1* and *TUBA1A*, were downregulated (**Figure 2**).

Pathway enrichment analysis of the DEGs was performed to interrogate signaling programs induced by SARS-CoV-2. Up-regulated genes were related to cytokine-cytokine receptor interaction, JAK-STAT signaling pathway, complement and coagulation cascades and other inflammatory pathways. Whereas, several metabolic pathways were negatively enriched, including ribosome, oxidative phosphorylation and ER protein processing, which suggests a global reduction in the production of proteins related to cellular energy production (**Figure 2**).

Transcriptome analysis of patients with mild, moderate and severe COVID-19

Next, we compared the transcriptomic profiles of patients with mild, moderate, and severe disease each separately with controls to determine if the above global signature, and the identified cytokine storm, hold similarly or varies by disease severity. Overall, we detected 2240 upregulated and 422 downregulated genes in mildly, 3447 upregulated and 302 downregulated genes in moderately, and 2691 upregulated and 215 downregulated genes in severely ill patient samples. Most of the immune response genes were significantly upregulated in all patient groups (**Supplementary Figure 1**).

Furthermore, pathway enrichment analysis showed modulation, to varying extents (**Figure 3**), of several overlapping pathways among mildly, moderately, and severely affected patients (**Supplementary Figure 1**). In addition to upregulation of immune-related response genes, there was consistent disruption of the ribosome pathway. The SARS-CoV-2 receptor ACE2 is an interferon-regulated gene and is upregulated in response to SARS-CoV-2 infection.[25] We also found that ACE2 expression was highest in patients with severe disease compared to patients with mild and moderate disease (**Figure 3**).

Comparison of functional enrichment among patients with mild, moderate, and severe COVID-19

To further delineate the differences among these groups, we compared the expression levels of genes and associated pathways (complement and coagulation cascades, cytokine-cytokine receptor interaction, JAK STAT signaling, TGF beta signaling pathway, platelets activation and ribosome pathway) between patients with mild, moderate, and severe disease. Gene expression in most regulated pathways was significantly higher (adj *p*-value < 0.05) in patients with severe disease compared to those with mild or moderate disease (**Figure 3**). For example, expression of certain interleukins (such as *IL11*, *IL12*, *IL19*, *IL34*), interleukin receptors (like *IL10RA*, *IL21R* and *IL11RA*), chemokines and tumour necrosis factor genes (such as *CXCL12*, *CXCL9*, *CCL25*, *CCL2*, *CCR5*, *CCR7*, *TNFSF9*, *TNFSF15*, *TNFRSF25*

and *TNFRSF9*) were significantly higher in patients with severe disease compared to patients with mild or moderate disease (adj *p*-value < 0.05; **Figure 3**). Furthermore, complement and coagulation cascades (e.g. *C2*, *C5*, *C6*, *F12* and *F8*) were activated to a higher extent in patients with severe disease. Finally, patients with severe disease had higher expression of *STAT4*, *STAT5A* and *STAT5B* which are important components of JAK-STAT pathway and play a critical role in the fate of T helper cells[26].

The highest fold change (> 2) in patients with severe disease was among the pro-inflammatory cytokines and chemokine receptors *CCR1* (CXCL8/CXCL6 receptor), *CCR6*, *CCR22*, *CCR25*, *IL3RA*, *IL11*, *IL19*, and *IL21RA*, the TGF-β signaling genes *BMP2*, *BMP7*, *PDGFA*, and *TNFSF11*, and the complement and platelet cascade genes *C6*, *C7*, *F2*, *F5*, *SERPINC1*, and *SERPIND1* (**Figure 4**). This distinction is likely due to viral infection resulting in release of inflammatory responses to a higher extent in patients with severe disease as compared to patients with mild or moderate disease.

Finally, we compared transcriptomic changes in male and female COVID-19 patients with mild and moderate outcomes. Several immune related genes (*CSF2*, *TNFSF11*, *TNFRSF11B*, *IL18R1*, *IFIT1B* and *C4BPA*) were upregulated in male patients compared to females (**Figure 5**).

Discussion

Transcriptomic analysis of nasopharyngeal tissues from patients with COVID-19 reveals a robust early induction of cytokine and immune-related profiles by SARS-CoV-2 infection. In addition, elevated levels of complement and coagulation cascades were seen. Our findings are similar to those recently observed by Butler et al.[27] in New York (United States) and Qin et al.[28] in Wuhan (China). The latter work reported that an increase in 'serum' cytokine and chemokine levels, as well as in neutrophil-lymphocyte-ratio (NLR) in SARS-CoV-2 infected patients is correlated with the severity of the disease and adverse outcomes, suggesting a role for hyper-inflammatory responses in COVID-19 pathogenesis. Similarly, inflammation-induced coagulation pathways, which can themselves be regulated by the complement system, are pivotal in controlling pathogenesis associated with infections.[29]

Our RNA work, using nasopharyngeal tissue, extends those findings to show that those pathways are more significantly altered in patients with severe disease outcomes.

Specifically, SARS-CoV-2 seems to induce expression of certain genes (**Figure 4**) to higher extents in patients with severe, compared to those with mild or moderate disease. This finding has direct prognostic and therapeutic implications. Identifying molecules targeting the pathways where those genes are involved might help ameliorate or avoid adverse clinical outcomes due to SARS-CoV-2. Moreover, this early expression signature can be used to predict the clinical course of disease, so that required early interventions can be implemented. However, larger datasets are needed to validate this expression signature and to show that it can reliably predict disease outcomes in the general population.

Certain molecular pathways require specific attention. In the majority of patients with COVID-19, respiratory failure is the primary cause of death and is often associated with uncontrolled inflammatory responses, edema, and lung fibrosis.[30] This lung fibrosis is mainly triggered by transforming growth factor-beta (TGF-β).[10] In our analysis, the TGF-β pathway was significantly upregulated in patients with severe disease suggesting they are more prone to respiratory failure. The PI3K/Akt and JAK/STAT signaling pathways are two other noteworthy pathways given that several cellular and immune responses, including host cell immune response to counteract viral infection [31], activation of interferon-stimulated genes (ISGs) [32], and activation of several regulatory and pro-inflammatory cytokines[33] converge on those two pathways.

Our gender-specific transcriptomic analysis revealed a few genes which were upregulated in male patients compared to females (**Figure 5**). It is highly likely that many other key factors, which are involved in sex-specific responses to SARS-CoV-2, were not detected in our study. It is possible that our cohort was not large enough to perform appropriate comparisons on a large number of patients adequately stratified for age, sex, and disease severity.

Taken together, our study highlights key molecular and functional pathways involved in COVID-19 pathogenesis and characterizes a specific early expression signature associated with severe disease outcomes due to SARS-CoV-2.

References

[1] Johns Hopkins Center for Systems Sciences and Engineering. COVID-19 Dashboard. https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6, 2020.

[2] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020, 395:497-506.

[3] Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS, China Medical Treatment Expert Group for C: Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020, 382:1708-1720.

[4] Uddin M, Mustafa F, Rizvi TA, Loney T, Suwaidi HA, Al-Marzouqi AHH, Eldin AK, Alsabeeha N, Adrian TE, Stefanini C, Nowotny N, Alsheikh-Ali A, Senok AC: SARS-CoV-2/COVID-19: Viral Genomics, Epidemiology, Vaccines, and Therapeutic Interventions. Viruses 2020, 12:526.

[5] vom Steeg LG, Klein SL: SeXX Matters in Infectious Disease Pathogenesis. PLoS Pathog 2016, 12:e1005374.

[6] Meng Y, Wu P, Lu W, Liu K, Ma K, Huang L, Cai J, Zhang H, Qin Y, Sun H, Ding W, Gui L, Wu P: Sex-specific clinical characteristics and prognosis of coronavirus disease-19 infection in Wuhan, China: A retrospective study of 168 severe patients. PLoS Pathog 2020, 16:e1008520.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.28.316604; this version posted September 28, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

7   Fehr AR, Channappanavar R, Perlman S: Middle East Respiratory Syndrome: Emergence of a Pathogenic Human Coronavirus. Annu Rev Med 2017, 68:387-399.

8   Newton AH, Cardani A, Braciale TJ: The host immune response in respiratory virus infection: balancing virus clearance and immunopathology. Semin Immunopathol 2016, 38:471-482.

9   de Wit E, van Doremalen N, Falzarano D, Munster VJ: SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 2016, 14:523-534.

10   Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, Guo D, Hu W, Yang J, Tang Z, Wu H, Lin Y, Zhang M, Zhang Q, Shi M, Liu Y, Zhou Y, Lan K, Chen Y: Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerg Microbes Infect 2020, 9:761-770.

11   Lieberman NAP, Peddu V, Xie H, Shrestha L, Huang ML, Mears MC, Cajimat MN, Bente DA, Shi PY, Bovier F, Roychoudhury P, Jerome KR, Moscona A, Porotto M, Greninger AL: In vivo antiviral host response to SARS-CoV-2 by viral load, sex, and age. bioRxiv 2020. doi: 10.1101/2020.06.22.165225

12   Abou Tayoun A, Loney T, Khansaheb H, Ramaswamy S, Harilal D, Deesi ZO, Varghese RM, Suwaidi HA, Alkhajeh A, AlDabal LM, Uddin M, Hamoudi R, Halwani R, Senok A, Hamid Q, Nowotny N, Alsheikh-Ali A. Genomic surveillance and phylogenetic analysis reveal multiple introductions of SARS-CoV-2 into a global travel hub in the Middle East. bioRxiv 2020. doi: 10.1101/2020.05.06.080606

13   Harilal D, Ramaswamy S, Loney T, Al Suwaidi H, Khansaheb H, Alkhaja A, Varghese R, Deesi Z, Nowotny N, Alsheikh-Ali A, Tayoun AA: SARS-CoV-2 Whole Genome Amplification and Sequencing for Effective Population-Based Surveillance and Control of Viral Transmission. Clin Chem 2020. doi: 10.1093/clinchem/hvaa187

14   Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, Melidou A, Neher RA, O'Toole A, Pereyaslov D, laboratories WHOERs, group GE, laboratories

WHOERs, group* GE: Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveill 2020, 25.

15 Andrews S: FastQC: a quality control tool for high throughput sequence data. 2010, Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

16 Ewels P, Magnusson M, Lundin S, Käller M: MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016, 32:3047-3048.

17 Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015, 12:357-360.

18 http://broadinstitute.github.io/picard/

19 Liao Y, Smyth GK, Shi W: featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2013, 30:923-930.

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

21 Yu G, Wang LG, Han Y, He QY: clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012, 16:284-287

22 https://software.broadinstitute.org/morpheus/

23 Goedhart J, Luijsterburg MS: VolcaNoseR – a web app for creating, exploring and sharing volcano plots. bioRxiv 2020. doi: 10.1101/2020.05.07.082263

24 GraphPad Prism version 8.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com

25 Ziegler CGK, Allon SJ, Nyquist SK, Mbano IM, Miao VN, Tzouanas CN, Cao Y, Yousif AS, Bals J, Hauser BM, Feldman J, Muus C, Wadsworth MH, 2nd, Kazer SW, Hughes TK, Doran B, Gatter GJ, Vukovic M, Taliaferro F, Mead BE, Guo Z, Wang JP, Gras D, Plaisant

M, Ansari M, Angelidis I, Adler H, Sucre JMS, Taylor CJ, Lin B, Waghray A, Mitsialis V, Dwyer DF, Buchheit KM, Boyce JA, Barrett NA, Laidlaw TM, Carroll SL, Colonna L, Tkachev V, Peterson CW, Yu A, Zheng HB, Gideon HP, Winchell CG, Lin PL, Bingle CD, Snapper SB, Kropski JA, Theis FJ, Schiller HB, Zaragosi LE, Barbry P, Leslie A, Kiem HP, Flynn JL, Fortune SM, Berger B, Finberg RW, Kean LS, Garber M, Schmidt AG, Lingwood D, Shalek AK, Ordovas-Montanes J, lung-network@humancellatlas.org HCALBNEa, Network HCALB: SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. Cell 2020, 181:1016-1035 e1019.

26 Seif F, Khoshmirsafa M, Aazami H, Mohsenzadegan M, Sedighi G, Bahar M: The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells. Cell Commun Signal 2017, 15:23.

27 Butler DJ, Mozsary C, Meydan C, Danko D, Foox J, Rosiene J, Shaiber A, Afshinnekoo E, MacKay M, Sedlazeck FJ, Ivanov NA, Sierra M, Pohle D, Zietz M, Gisladottir U, Ramlall V, Westover CD, Ryon K, Young B, Bhattacharya C, Ruggiero P, Langhorst BW, Tanner N, Gawrys J, Meleshko D, Xu D, Steel PAD, Shemesh AJ, Xiang J, Thierry-Mieg J, Thierry-Mieg D, Schwartz RE, Iftner A, Bezdan D, Sipley J, Cong L, Craney A, Velu P, Melnick AM, Hajirasouliha I, Horner SM, Iftner T, Salvatore M, Loda M, Westblade LF, Cushing M, Levy S, Wu S, Tatonetti N, Imielinski M, Rennert H, Mason CE: Shotgun Transcriptome and Isothermal Profiling of SARS-CoV-2 Infection Reveals Unique Host Responses, Viral Diversification, and Drug Interactions. bioRxiv 2020. doi: 10.1101/2020.04.20.048066

28 Qin C, Zhou L, Hu Z, Zhang S, Yang S, Tao Y, Xie C, Ma K, Shang K, Wang W, Tian D-S. Dysregulation of Immune Response in Patients with Coronavirus 2019 (COVID-19) in Wuhan, China. Clin Infect Dis. 2020, 71:762-768.

29 Ramlall V, Thangaraj PM, Meydan C, Foox J, Butler D, Kim J, May B, De Freitas JK, Glicksberg BS, Mason CE, Tatonetti NP, Shapira SD: Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. Nature Medicine 2020. doi: 10.1038/s41591-020-1021-2 .

[30] Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Liu S, Zhao P, Liu H, Zhu L, Tai Y, Bai C, Gao T, Song J, Xia P, Dong J, Zhao J, Wang F-S: Pathological findings of COVID-19 associated with acute respiratory distress syndrome. The Lancet Respiratory Medicine 2020, 8:420-422.

[31] Diehl N, Schaal H: Make yourself at home: viral hijacking of the PI3K/Akt signaling pathway. Viruses 2013, 5:3192-3212.

[32] Ezell SA, Polytarchou C, Hatziapostolou M, Guo A, Sanidas I, Bihani T, Comb MJ, Sourvinos G, Tsichlis PN: The protein kinase Akt1 regulates the interferon response through phosphorylation of the transcriptional repressor EMSY. Proc Natl Acad Sci U S A 2012, 109:E613-621.

[33] Villarino AV, Kanno Y, Ferdinand JR, O'Shea JJ: Mechanisms of Jak/STAT signaling in immunity and disease. J Immunol 2015, 194:21-27.

Figure Legends

Figure 1. Schematic workflow for Transcriptomic analysis. Sequencing data underwent pre-processing which includes primary QC and read mapping, followed by differenticially expressed genes (DEG) analysis and downstream pathway enrichment analysis and visualization.

Figure 2. *Left*, Dotplot visualization of enriched Pathway terms in all COVID-19 patients. The color of the dots represents adj *p*-value for each enriched pathway, and size represents the percentage of genes enriched in the total gene set. *Right*, Volcano plot representing upregulated and downregulated genes. X-axis represents $\log_2$ fold change of genes and Y-axis represents $-\log_{10}$ *P*-value in differentially expressed gene (DEG) analysis.

Figure 3. Heat map and Violin plots of pathways in patients with mild, moderate or severe disease. Pathways are: a) Complement and coagulation cascades, Cytokine-cytokine receptor interaction, JAK-STAT signaling pathway, d) TGF-beta signalling, Platelets activation, and Ribosome. Heat map depicts the $\log_2$ fold change of differentially expressed gene (DEGs) of

COVID-19 patients compared with controls. Genes included have a $\log_2$ fold change of more than 1 and a *p*-adjusted value of less than 0.05. For every gene in a given pathway, the average counts per million (CPM) was calculated in cases with mild, moderate and severe disease and these CPM values were plotted (Violin plots) indicating median and quartiles as well as minima and maxima bounds. *Expression of *ACE2* and *TMPRSS2*.

Figure 4: Scatter plot of fold change gene expression differences between patients with mild/moderate and severe outcomes. Fold change of mild/moderate was averaged and subtracted from severe fold change when compared to controls.

Figure 5: Heat map (top) and Violin plots (bottom) of differentially expressed genes in males and females. Heat map depicts the $\log_2$ fold change of differentially expressed gene (DEGs) of COVID-19 patients compared with controls. Genes included have a $\log_2$ (fold change) of more than 1 and adj *p*-value < 0.05. Violin plots represent median and 75th quartiles as well as minima and maxima bounds of CPM (counts per million) of highlighted differentially regulated genes.

Supplementary figure 1: Dotplot visualization of upregulated (a), and downregulated pathways (b) in patients with mild, moderate and severe disease. The color of the dots represents adj *p*-value for each enriched pathways, and size represents the percentage of genes enriched in the total gene set. Volcano plots for upregulated and downregulated genes in patient with mild (c), moderate (d), and severe (e) outcomes. X-axis represents $\log_2$ fold change of genes and y axis represents $-\log_{10}$ *P*-value in DEG analysis.

Table 1. Clinical Characteristics of COVID-19 patients in this study

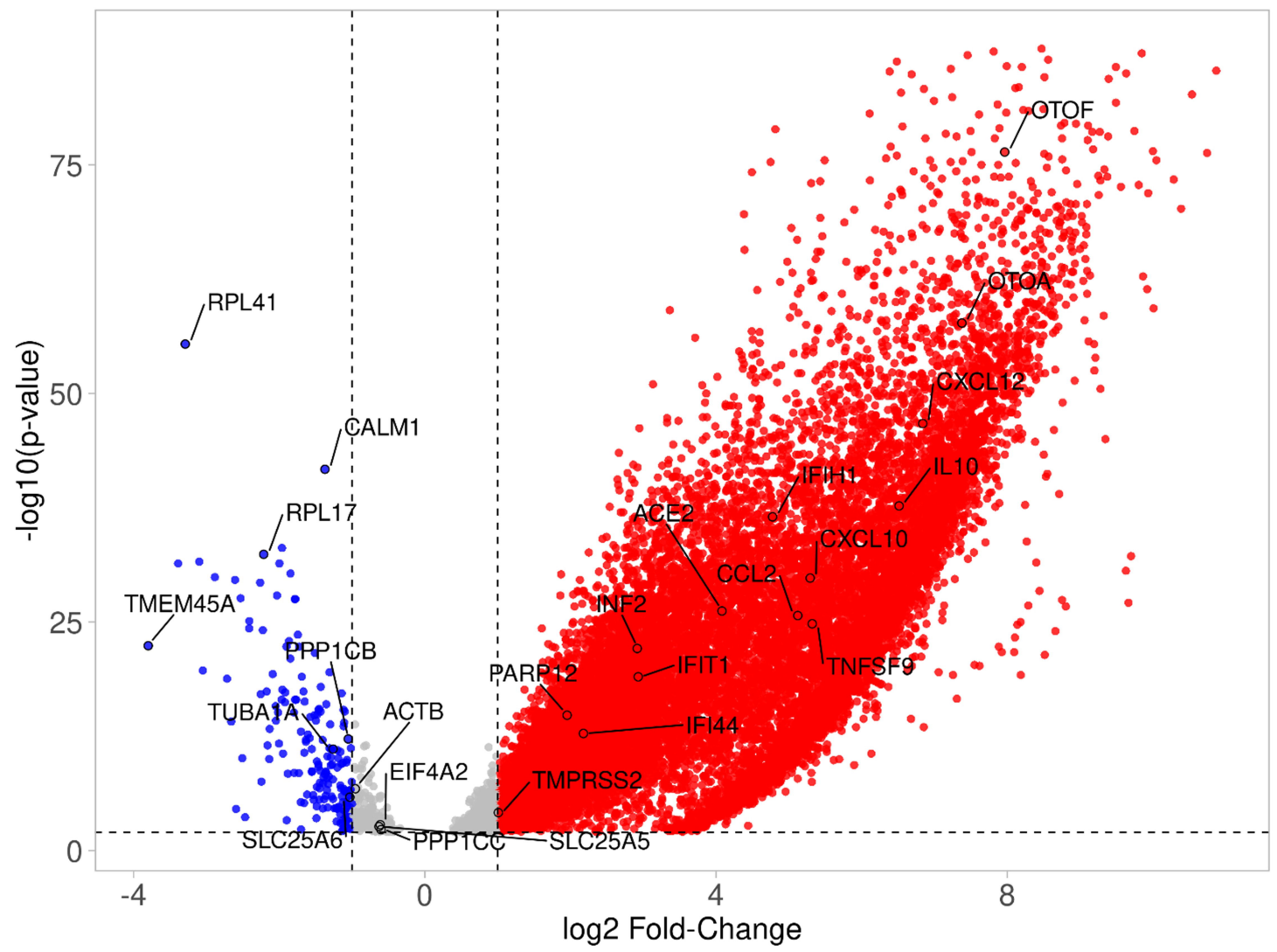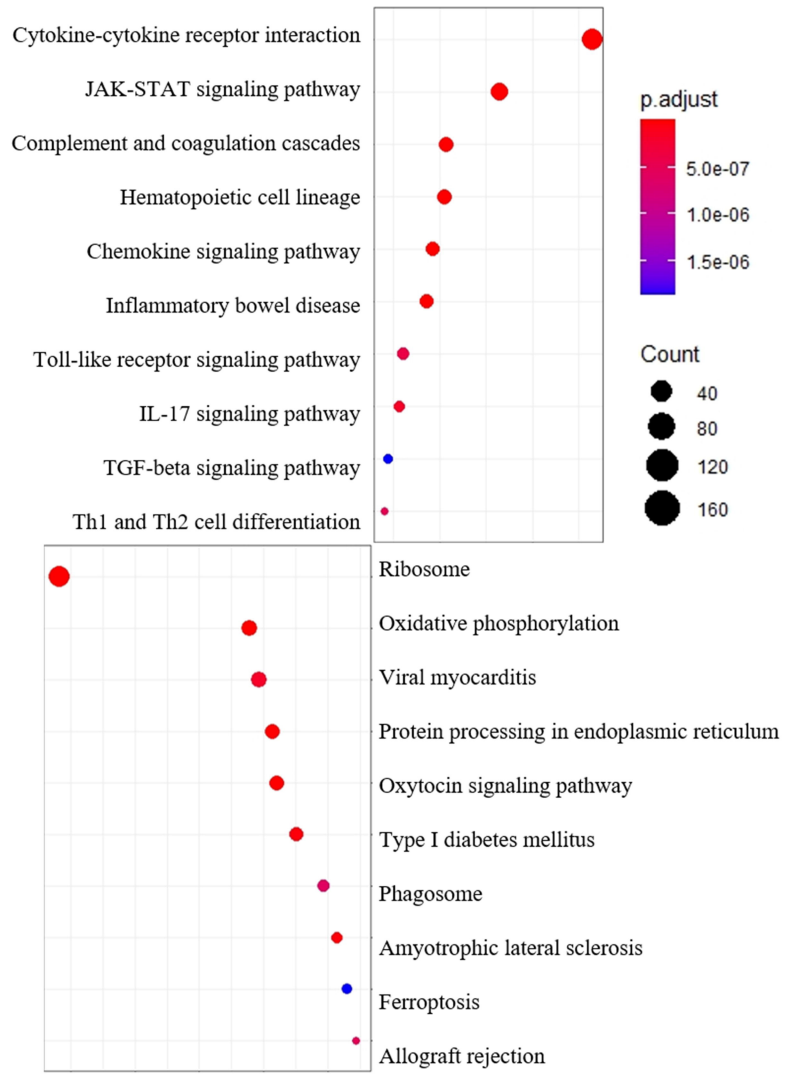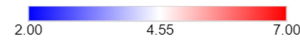| Characteristic	All Patients (N=50)	Stratified by Disease Severity		
		Asymptomatic/Mild (N=37)	Moderate (N=10)	Severe/Critical (N=3)
Age Mean (SD), yr	40.92 (16.03)	36.4 (14.04)	49.28 (14.35)	68 (3.56)
Female Sex - no./total no. (%)	18/50 (36%)	16/37 (43.24%)	1/10 (10%)	1/3 (33.33%)
Body Mass (kg) - Mean (SD)	66.2 (30.84)	63.4 (28.65)	73.07 (41.83)	77.93 (8.29)
Current Smoker - no./total no. (%)	0/50 (0%)	0/37 (0%)	0/10 (0%)	0/3 (0%)

Coexisting Disorder - no./total no. (%)				
Chronic Cardiac Disease (not Hypertension)	3/50 (6%)	1/37 (2.7%)	2/10 (20%)	0/3 (0%)
Hypertension	9/50 (18%)	5/37 (13.51%)	3/10 (30%)	1/3 (33.33%)
Chronic Pulmonary Disease	0/50 (0%)	0/37 (0%)	0/10 (0%)	0/3 (0%)
Asthma	0/50 (0%)	0/37 (0%)	0/10 (0%)	0/3 (0%)
Chronic kidney disease	1/50 (2%)	0/37 (0%)	0/10 (0%)	1/3 (33.33%)
Chronic liver disease	0/50 (0%)	0/37 (0%)	0/10 (0%)	0/3 (0%)
Chronic Neurological Disorder	0/50 (0%)	0/37 (0%)	0/10 (0%)	0/3 (0%)
Diabetes Mellitus	8/50 (16%)	3/37 (8.1%)	3/10 (30%)	2/3 (66.66%)
Malignant Neoplasms	0/50 (0%)	0/37 (0%)	0/10 (0%)	0/3 (0%)

1. Primary QC and read mapping

Reads → **Quality control** → **Reads mapped to human rRNA**

a) FastQC v0.11.8

b) MultiQC v1.6

hisat v2-2.1.0

Coverage QC ← **Unaligned reads mapped to SARS-CoV-2 genome** ← **Reads mapped to Hg19 genome**

Picard v2.18.17

BWA v0.7.17

hisat v2-2.1.0
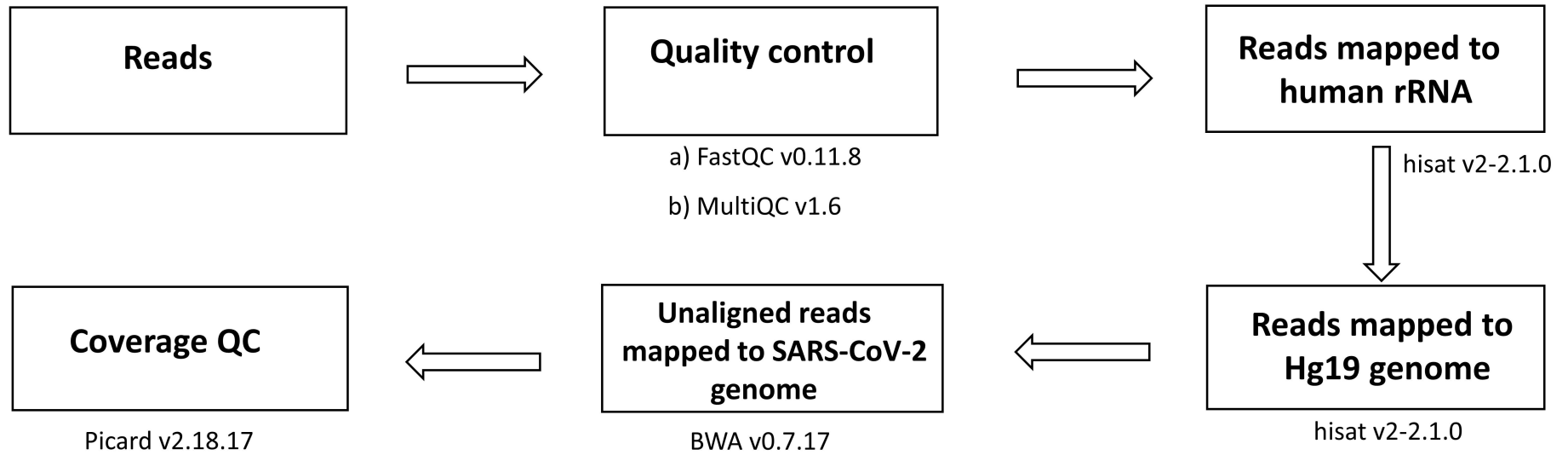
2. Expression profiling and pathway enrichment analysis
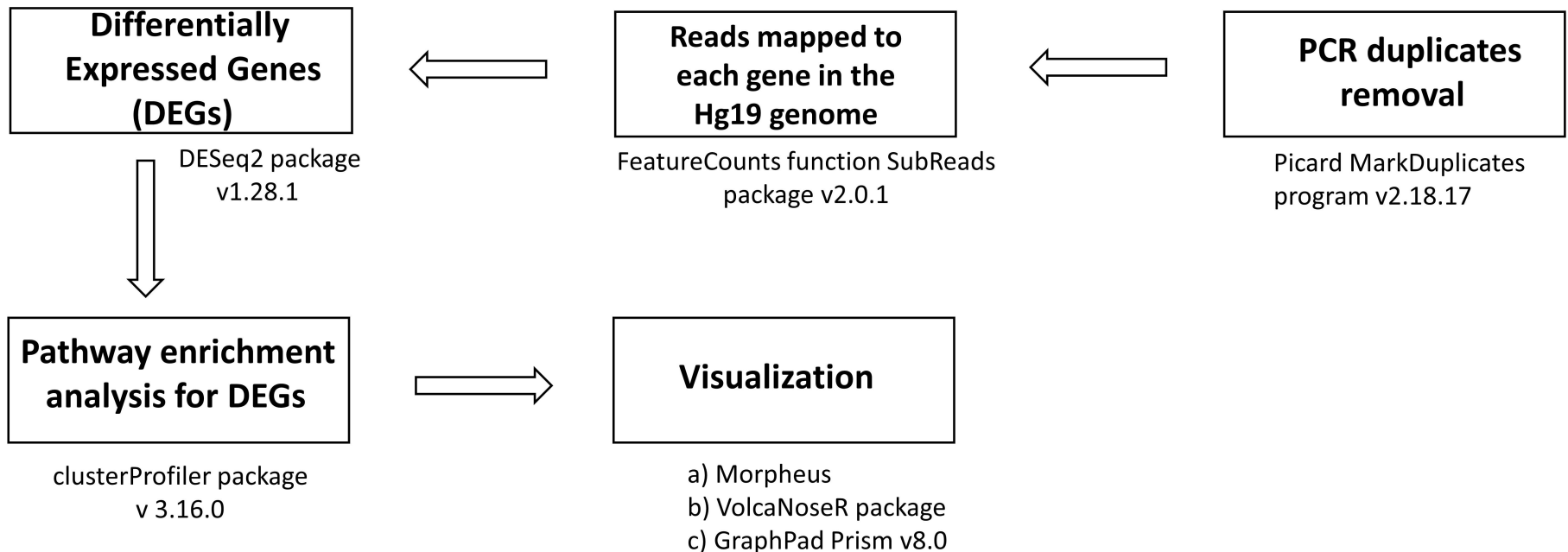
Differentially Expressed Genes (DEGs) ← **Reads mapped to each gene in the Hg19 genome** ← **PCR duplicates removal**

DESeq2 package v1.28.1

FeatureCounts function SubReads package v2.0.1

Picard MarkDuplicates program v2.18.17

Pathway enrichment analysis for DEGs → **Visualization**

clusterProfiler package v 3.16.0

a) Morpheus

b) VolcaNoseR package

c) GraphPad Prism v8.0
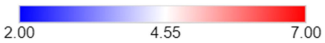
Complement and coagulation cascades
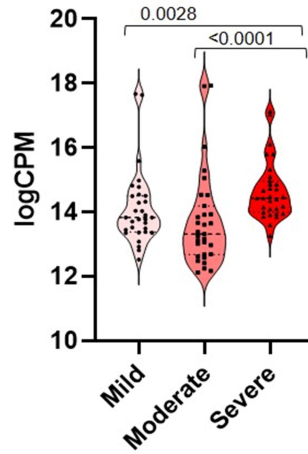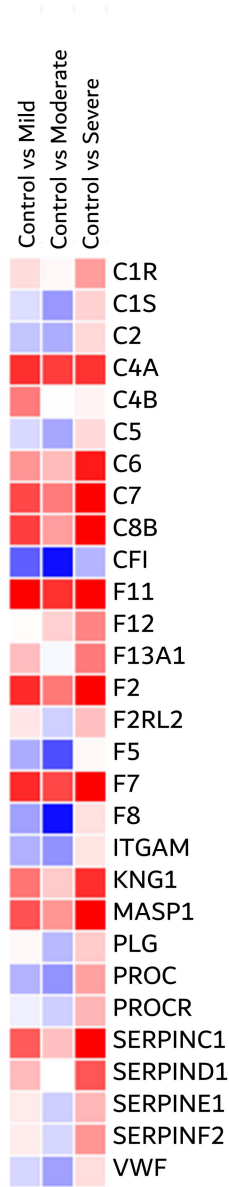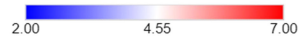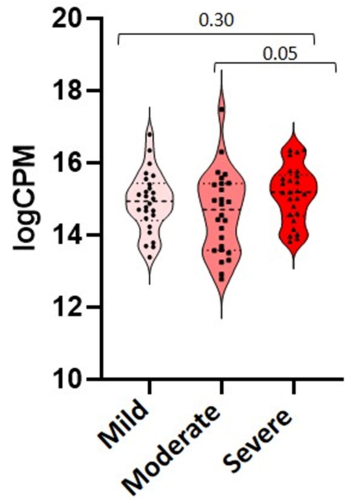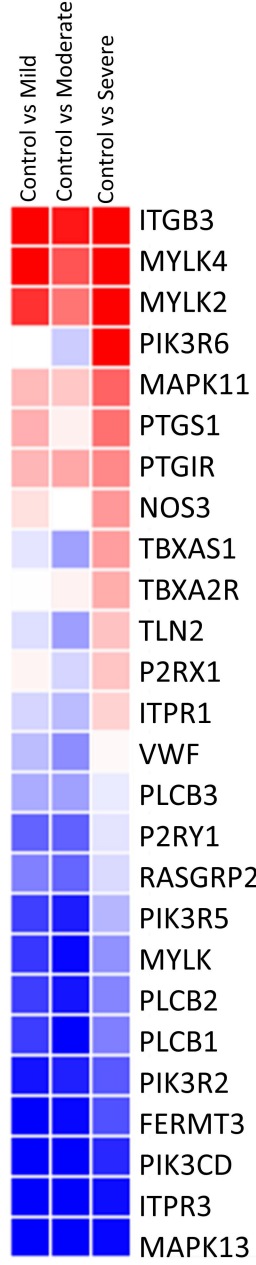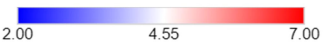
Cytokine-cytokine receptor interaction
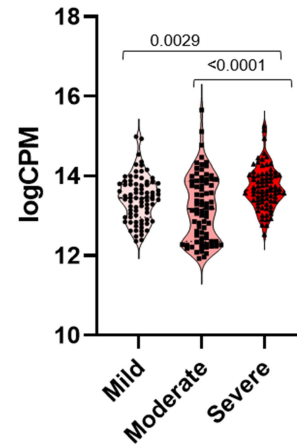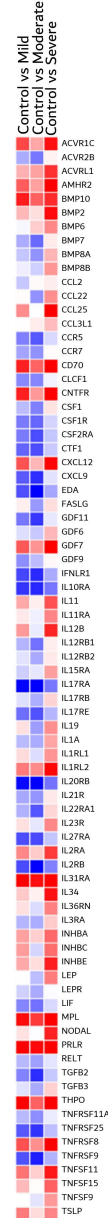
JAK-STAT signaling pathway
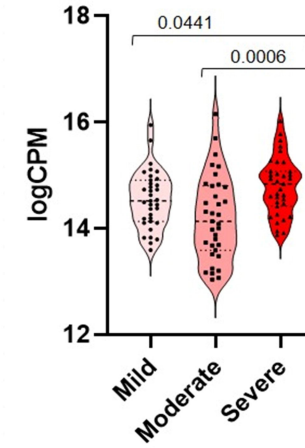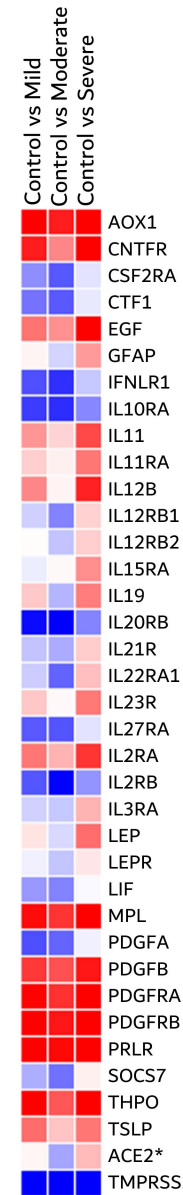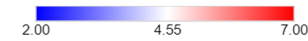
TGF-beta signaling pathway
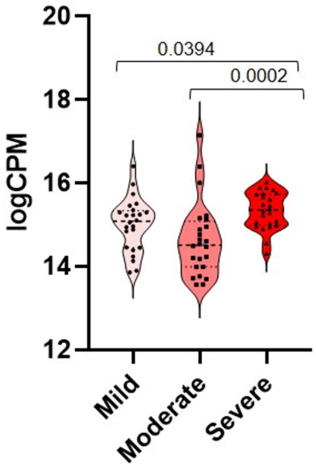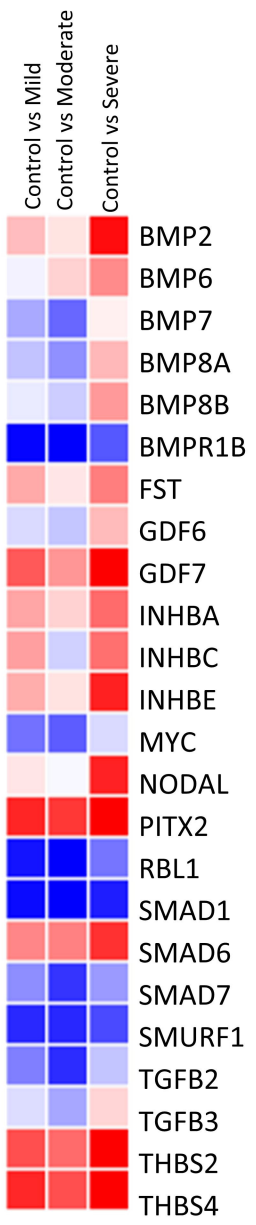
Platelets activation

Ribosomes