

---

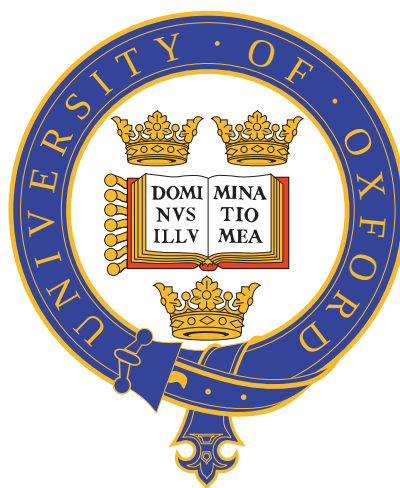
---

# An integrated immune -omic approach to unravel the sepsis response

---

---

Andrew J. Kwok



Wadham College

UNIVERSITY OF OXFORD

A thesis submitted in partial fulfilment of  
the requirements for the degree of Doctor of  
Philosophy

TRINITY TERM, 2022



## **An integrated immune -omic approach to unravel the sepsis response**

Andrew J. Kwok, Wadham College, Trinity Term 2022

*A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy of the University of Oxford*

Sepsis is defined as life-threatening organ dysfunction due to a dysregulated host response to infection. There is major unmet clinical need given the burden of all-cause sepsis accounting for 20% of global deaths, in addition to the COVID-19 pandemic where SARS-CoV-2 can lead to viral sepsis. Multiple dynamic host pathophysiological mechanisms exist but are aggregated under a single umbrella syndrome description, leading to substantial patient heterogeneity. This thesis examines the sepsis response through a lens of immune multi-omics, focusing on single-cell transcriptomics with additional proteomic profiling.

Data are presented for novel sepsis patient cohorts recruited under different clinical emphases, including patients with COVID-19 (n=78), influenza (n=12), and all-cause sepsis (n=63), and reanalysis of existing cohorts (n=667). The host response to severe infection is explored via bulk and single-cell transcriptomics (scRNA-seq), and proteomics either simultaneous to scRNA-seq i.e. single-cell multi-omics, or in isolation with CyTOF. Comparisons are made between severe infection/sepsis comparator groups, within groups across transcriptomic subphenotypes, and with healthy and post-cardiac surgery control patients.

The host transcriptomic response showed substantial commonality between COVID-19 viral sepsis and all-cause sepsis although immunoglobulin pathways and the interferon response were elevated in COVID-19. The sepsis response signature (SRS) transcriptomic subphenotyping proved consistent in both cohorts, suggesting its utility as a unifying framework to view the host response to severe infection. At single-cell resolution, transcriptomics and cell surface proteomics of peripheral blood mononuclear cells revealed common and distinct aspects of the response to severe infection across cell types such as classical monocytes and CD4<sup>+</sup> T cells, with a specific COVID-19 signature uncovered in the form of AP-1 pathway elevation. The basis of SRS was then dissected, revealing that SRS derived mostly from neutrophils with a minor contribution from PBMCs. Sepsis neutrophils displayed immunosuppressive properties, in particular SRS1 neutrophils, suggesting a potential role in sepsis pathogenesis.

Collectively, the work presented in this thesis advances our understanding of the host response to severe infection in terms of the cellular and molecular bases of the sepsis response. This biological understanding highlights potential therapeutic opportunities as we move toward a personalized infectious disease medicine in sepsis and beyond.

## ACKNOWLEDGEMENTS

---

First, I would like to thank my supervisors Julian Knight, John Todd and Justin Whalley. Julian has been immensely supportive throughout the past four years of the various ideas that I've been keen to explore. This has enabled me to pursue a wide range of projects and expand my thinking in multiple directions. I am grateful for his patience and his ability to always find time to discuss things big and small despite running such a large lab, and giving me the space to develop my scientific thinking. John has provided me with key nuggets of wisdom and advice at timely moments throughout. I am thankful to Justin for his bioinformatics supervision, especially in the beginning when I started with no bioinformatic expertise at all.

I would also like to thank members of the Knight group. Fellow PhD students Patrick Maclean and Yuxin Mi have provided critical discussion, and more importantly, helped with ensuring sanity throughout the PhD. I am grateful to Andrew Brown for his help especially in the earlier stages in helping me get up onto my own two feet in the lab. Giuseppe Scozzafava and Alice Trickett have been constant sources of supports. Alice has been particularly inspiring in the lab with her work ethic and efficiency, and has been a fantastic lab partner to work with and troubleshoot pesky lab experiments. Ricardo Ferreira and Madeleine Smee also deserve my gratitude for their work together in the lab, as well as members of the COMBAT consortium who made the intense work during lockdown much more bearable. I am thankful to Giorgio Napolitani, Katie Burnham, Eddie Cano-Gamez and Emma Davenport who have been more than just collaborators, but rather, great mentors from whom I have learnt so much from. I also thank Sally Beer and her research team for their incredible enthusiasm in helping with study set up and patient recruitment.

I am also enormously grateful to members of the Oxford Genomics Centre, especially Moustafa Attar for his generosity and boundless wisdom and knowledge when it comes to all things single-cell, as well as Fabiola Curion for her guidance in single-cell bioinformatics. Angela Lee has helped ensure that our projects have run smoothly and that the work we have produced together has always been of the highest quality.

I would like to acknowledge the support of my friends inside and outside of Oxford throughout the past few years, especially Francesco Moiraghi, for the thought-provoking discussions about science, philosophy of science, and all things beyond. Finally, I would not have succeeded in completing this journey without the constant support of my family, my parents and Kate Yoon in particular. They have listened to all my ups and downs and lived through the roller coasters of the past four years with me, giving me steady encouragement whenever there have been tough times.

## DECLARATIONS

---

I declare that, unless otherwise stated, all work presented in this thesis is my own. Several aspects of the study relied upon collaboration where part of the work was conducted with or by others.

**Bulk RNA-seq:** Library preparations were performed by the Oxford Genomics Centre (Wellcome Centre for Human Genetics Core Genomics facilities). Analysis was performed in conjunction with Katie Burnham and Emma Davenport from the Wellcome Sanger Institute.

**COMBAT influenza patient recruitment:** Influenza samples were provided by collaborators from London St George's University Hospitals NHS Foundation Trust, Guy's and St Thomas' NHS Foundation Trust and King's College Hospital NHS Foundation Trust. Details are given in the materials and methods.

**COMBAT CITE-seq:** The CITE-seq experiment and sequencing library preparations were carried out together with Moustafa Attar and Hubert Slawinski from the Oxford Genomics Centre. Analysis of CITE-seq data was performed by the COMBAT consortium as a whole, of which I was a key member of in all steps. Fabiola Curion, Santiago Revalle, Steven Sansom, Kathrin Jensen and Maria Gomez Vazquez assisted in the pre-processing and quality control. Calliope Dendrou and Steven Sansom led the clustering and annotation teams which I was part of, while Charlotte Rich-Griffin and Lucy Garner performed the differential abundance analysis under the supervision of Calliope Dendrou, with input from the wider CITE-seq team. I was also part of the WGCNA team which was also led by Steven Sansom along with Fabiola Curion, Charlotte Rich-Griffin and Lucy Garner.

**Neutrophil functional assays:** I designed the neutrophil functional assays, which were then carried out by Alice Trickett and Madeleine Smee from the Knight lab. I was involved in the analysis strategy and supervision of these experiments, while formal analysis was carried out by Alice Trickett and Madeleine Smee.

**CyTOF data:** The CyTOF data from the 53 samples in the third results chapter were generated, debarcoded and doublet cleaned by David Ahern from the Kennedy Institute of Rheumatology as part of the COMBAT project, with supervision from Irina Udalova and Claudia Monaco. I performed the analysis for the dataset with assistance in myeloid cell cluster annotation from Yasemin-Xiomara Zurke, Irina Udalova, Claudia Monaco, David Ahern and Lihui Wang.

## SUBMITTED ABSTRACTS

---

### **Identification of deleterious neutrophil states and altered granulopoiesis in sepsis**

Poster: Association of Physicians of Great Britain and Ireland Annual Meeting (2022). Best poster award.

**AJ Kwok**, A Allcock, RC Ferreira, M Smee, E Cano-Gamez, EMROx, S McKechnie, CJ Hinds, JA Todd, JC Knight

### **Immature granulocytes and heightened emergency granulopoiesis underlie a deleterious sepsis endotype**

Wellcome Connecting Science: Single Cell Biology (2022)

**AJ Kwok**, RC Ferreira, E Cano-Gamez, EMROx, S McKechnie, CJ Hinds, JA Todd, JC Knight

## ASSOCIATED PUBLICATIONS

---

**An integrated multi-omics blood atlas reveals signatures and drivers of severe COVID-19**

Cell (2022), Volume 185, Issue 5, Pages 916-938.e58

COMBAT Consortium

- Lead junior author in consortium

**An immune dysfunction score for stratification of patients with acute infection based on whole blood gene expression**

medRXiv (2022), doi: <https://doi.org/10.1101/2022.03.17.22272427>

E Cano-Gamez, KL Burnham, C Goh, ZH Malick, **AJ Kwok**, DA Smith, H Peters-Sengers, D Antcliffe, GAIN Investigators, S McKechnie, BP Scicluna, T van der Poll, AC Gordon, CJ Hinds, EE Davenport, JC Knight

**Identification of deleterious neutrophil states and altered granulopoiesis in sepsis**

medRXiv (2022), doi: <https://doi.org/10.1101/2022.03.22.22272723>

**AJ Kwok**, A Allcock, RC Ferreira, M Smee, E Cano-Gamez, KL Burnham, YX Zurke, EMROx, S McKechnie, C Monaco, I Udaloova, CJ Hinds, EE Davenport, JA Todd, JC Knight

# CONTENTS

---

|                                                                                                                 |            |
|-----------------------------------------------------------------------------------------------------------------|------------|
| <b>Abstract</b>                                                                                                 | <b>i</b>   |
| <b>Acknowledgements</b>                                                                                         | <b>ii</b>  |
| <b>Declarations</b>                                                                                             | <b>iii</b> |
| <b>Submitted Abstracts</b>                                                                                      | <b>iv</b>  |
| <b>Associated Publications</b>                                                                                  | <b>v</b>   |
| <b>Contents</b>                                                                                                 | <b>vi</b>  |
| <b>List of Figures</b>                                                                                          | <b>ix</b>  |
| <b>List of Tables</b>                                                                                           | <b>x</b>   |
| <b>Abbreviations</b>                                                                                            | <b>xi</b>  |
| <b>1 Introduction</b>                                                                                           | <b>1</b>   |
| 1.1 The clinical syndrome of sepsis . . . . .                                                                   | 2          |
| 1.2 SARS-CoV-2 and COVID-19 . . . . .                                                                           | 8          |
| 1.3 The immune system . . . . .                                                                                 | 12         |
| 1.4 Prospective cohort studies – the GAINs and Sepsis Immunomics<br>Studies and the COMBAT consortium . . . . . | 18         |
| 1.5 Functional genomics in sepsis and COVID-19 . . . . .                                                        | 21         |
| 1.6 Specific aims and objectives . . . . .                                                                      | 30         |
| <b>2 Materials and Methods</b>                                                                                  | <b>34</b>  |
| 2.1 Patient cohorts in chapters 3 and 4 . . . . .                                                               | 34         |
| 2.2 Whole blood bulk RNA-seq . . . . .                                                                          | 37         |
| 2.3 Peripheral blood mononuclear cell (PBMC) CITE-seq . . . . .                                                 | 39         |
| 2.4 Patient cohorts for chapter 5 . . . . .                                                                     | 46         |
| 2.5 Whole blood scRNA-seq and Abseq . . . . .                                                                   | 47         |
| 2.6 Neutrophil functional assays . . . . .                                                                      | 53         |
| 2.7 Chapter 5 cohort 2/3 analyses . . . . .                                                                     | 55         |
| <b>3 The Bulk Transcriptomic Response to All-cause and COVID-19 Sepsis</b>                                      | <b>61</b>  |
| 3.1 Introduction . . . . .                                                                                      | 61         |
| 3.2 Results . . . . .                                                                                           | 66         |
| 3.3 Discussion . . . . .                                                                                        | 98         |
| 3.4 Conclusions . . . . .                                                                                       | 107        |
| <b>4 A Large Scale Single-cell Multi-omic PBMC Atlas of Sepsis, Severe<br/>COVID-19 and Influenza</b>           | <b>108</b> |
| 4.1 Introduction . . . . .                                                                                      | 108        |
| 4.2 Results . . . . .                                                                                           | 116        |

## CONTENTS

---

|          |                                                                          |            |
|----------|--------------------------------------------------------------------------|------------|
| 4.3      | Discussion . . . . .                                                     | 146        |
| 4.4      | Conclusions . . . . .                                                    | 151        |
| <b>5</b> | <b>Whole Blood Single-cell Multi-omics for Sepsis Subphenotypes</b>      | <b>153</b> |
| 5.1      | Introduction . . . . .                                                   | 153        |
| 5.2      | Results . . . . .                                                        | 161        |
| 5.3      | Discussion . . . . .                                                     | 198        |
| 5.4      | Conclusions . . . . .                                                    | 205        |
| <b>6</b> | <b>General Discussion</b>                                                | <b>206</b> |
| 6.1      | Multi-omics . . . . .                                                    | 207        |
| 6.2      | Tissues of interest in severe infection . . . . .                        | 212        |
| 6.3      | Tissues of interest in the context of legacy effects of severe infection | 214        |
| 6.4      | Animal models for SRS . . . . .                                          | 216        |
| 6.5      | Translational opportunities . . . . .                                    | 217        |
| 6.6      | Limitations and future work . . . . .                                    | 219        |
| 6.7      | Conclusion . . . . .                                                     | 222        |
|          | <b>Appendices</b>                                                        | <b>224</b> |
| <b>A</b> | <b>Appendix to chapter 2</b>                                             | <b>224</b> |
| <b>B</b> | <b>Appendix to chapter 3</b>                                             | <b>233</b> |
| <b>C</b> | <b>Appendix to chapter 4</b>                                             | <b>238</b> |
| <b>D</b> | <b>Appendix to chapter 5</b>                                             | <b>249</b> |
|          | <b>Bibliography</b>                                                      | <b>267</b> |

## LIST OF FIGURES

|      |                                                                |     |
|------|----------------------------------------------------------------|-----|
| 3.1  | Bulk RNA-seq PCA . . . . .                                     | 69  |
| 3.2  | PCA covariate associations . . . . .                           | 71  |
| 3.3  | PCA after age/sex adjustment . . . . .                         | 72  |
| 3.4  | Hierarchical clustering analysis . . . . .                     | 73  |
| 3.5  | Unsupervised analysis of non-COVID-19 sepsis samples . . . . . | 75  |
| 3.6  | SRS sensitivity analysis . . . . .                             | 77  |
| 3.7  | Non-COVID-19 SRS DGE . . . . .                                 | 79  |
| 3.8  | SRS expression changes vs. original derivation . . . . .       | 80  |
| 3.9  | SRS mortality analysis . . . . .                               | 81  |
| 3.10 | COVID-19 SRS DGE . . . . .                                     | 83  |
| 3.11 | COVID-19 SRS evaluation . . . . .                              | 84  |
| 3.12 | Sepsis-HV DGE . . . . .                                        | 86  |
| 3.13 | Sepsis-HV DGE overlaps . . . . .                               | 87  |
| 3.14 | Sepsis-HV pathway analysis . . . . .                           | 88  |
| 3.15 | SRS DGE . . . . .                                              | 89  |
| 3.16 | SRS DGE overlaps . . . . .                                     | 90  |
| 3.17 | SRS upregulated pathways . . . . .                             | 91  |
| 3.18 | SRS downregulated pathways . . . . .                           | 92  |
| 3.19 | All-cause vs. COVID-19 sepsis DGE . . . . .                    | 94  |
| 3.20 | All-cause vs. COVID-19 sepsis DGE overlaps . . . . .           | 95  |
| 3.21 | All-cause vs. COVID-19 sepsis upregulated pathways . . . . .   | 96  |
| 3.22 | All-cause vs. COVID-19 sepsis downregulated pathways . . . . . | 97  |
| 3.23 | Septic vs. non-septic COVID-19 DGE . . . . .                   | 98  |
| 3.24 | Septic vs. non-septic COVID-19 DGE overlaps . . . . .          | 99  |
| 3.25 | Septic vs. non-septic COVID-19 pathway analysis . . . . .      | 100 |
| 4.1  | CITE-seq experimental design . . . . .                         | 118 |
| 4.2  | CITE-seq experimental workflow . . . . .                       | 119 |
| 4.3  | Cell numbers per sample post QC . . . . .                      | 119 |
| 4.4  | ADT clustering . . . . .                                       | 121 |
| 4.5  | MNP subsetting and clustering strategy . . . . .               | 122 |
| 4.6  | MNP annotations . . . . .                                      | 123 |
| 4.7  | MNP CD14 expression . . . . .                                  | 124 |
| 4.8  | T/NK cell subsetting and clustering strategy . . . . .         | 126 |
| 4.9  | CD4+ T cell annotations . . . . .                              | 127 |
| 4.10 | CD8+ T cell annotations . . . . .                              | 128 |
| 4.11 | NK cell annotations . . . . .                                  | 130 |
| 4.12 | B cell subsetting and clustering strategy . . . . .            | 131 |
| 4.13 | B cell annotations . . . . .                                   | 132 |
| 4.14 | HSC annotations . . . . .                                      | 133 |
| 4.15 | PBMC minor subset differential abundance . . . . .             | 136 |
| 4.16 | PBMC cluster differential abundance . . . . .                  | 137 |
| 4.17 | WGCNA soft thresholding . . . . .                              | 140 |
| 4.18 | WGCNA module construction . . . . .                            | 141 |
| 4.19 | WGCNA modules . . . . .                                        | 142 |



## LIST OF FIGURES

---

|                                                                       |     |
|-----------------------------------------------------------------------|-----|
| 4.20 WGCNA AP-1 modules . . . . .                                     | 145 |
| 4.21 WGCNA modules with disease condition related expression patterns | 146 |
| 5.1 Chapter 5 strategy . . . . .                                      | 162 |
| 5.2 Cohort 1 Abseq lineage markers . . . . .                          | 165 |
| 5.3 Cohort 1 broad annotation . . . . .                               | 166 |
| 5.4 Cohort 1 immune hallmarks of sepsis and systemic inflammation .   | 167 |
| 5.5 Cohort 1 fine clustering approach . . . . .                       | 168 |
| 5.6 Cohort 1 immature neutrophil markers . . . . .                    | 169 |
| 5.7 Cohort 1 neutrophil trajectory analysis . . . . .                 | 169 |
| 5.8 Cohort 1 sepsis-HV differential abundance . . . . .               | 170 |
| 5.9 Cohort 1 cardiac surgery-HV differential abundance . . . . .      | 171 |
| 5.10 Cohort 1 sepsis-cardiac surgery differential abundance . . . . . | 171 |
| 5.11 Sepsis neutrophil-CD4+ T cell co-culture . . . . .               | 173 |
| 5.12 Sepsis neutrophil-CD4+ T cell co-culture reversal . . . . .      | 174 |
| 5.13 Cohort 1 SRS . . . . .                                           | 176 |
| 5.14 Cohort 1 SRS differential abundance . . . . .                    | 176 |
| 5.15 Cohort 1 neutrophil SRS differential gene expression . . . . .   | 178 |
| 5.16 Cohort 1 MNC SRS differential gene expression . . . . .          | 179 |
| 5.17 Sepsis neutrophil functional assays . . . . .                    | 180 |
| 5.18 Cohort 2 bulk RNA-seq WGCNA . . . . .                            | 181 |
| 5.19 Cohort 2 CyTOF neutrophil clustering . . . . .                   | 182 |
| 5.20 Cohort 2 CyTOF neutrophils . . . . .                             | 183 |
| 5.21 Cohort 2 CyTOF neutrophil differential abundance . . . . .       | 184 |
| 5.22 Cohort 2 CyTOF neutrophil trajectory analysis . . . . .          | 185 |
| 5.23 Cohort 2 CyTOF PCA . . . . .                                     | 187 |
| 5.24 Cohort 2 multi-omic factor analysis setup . . . . .              | 188 |
| 5.25 Cohort 2 multi-omic factor analysis by SRS . . . . .             | 188 |
| 5.26 Cohort 2 multi-omic factor analysis loadings . . . . .           | 190 |
| 5.27 Cohort 2 IL1R2+ immature neutrophil validation . . . . .         | 191 |
| 5.28 Cohort 3 cell proportion deconvolution . . . . .                 | 192 |
| 5.29 Cohort 1 SRS neutrophil cNMF . . . . .                           | 194 |
| 5.30 Cohort 1 neutrophil gene programs . . . . .                      | 195 |
| 5.31 Cohort 1 neutrophil gene program TF prediction . . . . .         | 196 |
| 5.32 Cohort 2 myelopoietic cytokines . . . . .                        | 197 |
| 5.33 SRS neutrophil NETosis . . . . .                                 | 198 |
| A.1 Bulk RNA-seq QC . . . . .                                         | 224 |
| A.2 CITE-seq QC . . . . .                                             | 230 |
| A.3 scRNA-seq + Abseq QC . . . . .                                    | 232 |
| C.1 CITE-seq minor subset proportions . . . . .                       | 238 |
| D.1 scRNA-seq + Abseq clustering evaluation 1 . . . . .               | 250 |
| D.2 scRNA-seq + Abseq clustering evaluation 2 . . . . .               | 251 |
| D.3 scRNA-seq + Abseq clustering evaluation 3 . . . . .               | 252 |

## LIST OF TABLES

|     |                                                                                    |     |
|-----|------------------------------------------------------------------------------------|-----|
| 3.1 | Clinical characteristics of bulk RNA-seq cohort . . . . .                          | 68  |
| 3.2 | SRS assignments . . . . .                                                          | 76  |
| 3.3 | COVID-19 SRS assignment . . . . .                                                  | 82  |
| 4.1 | Clinical characteristics of CITE-seq cohort . . . . .                              | 117 |
| 5.1 | Clinical characteristics of cohort 1 . . . . .                                     | 163 |
| 5.2 | Cohort 1 sepsis patient microbiology . . . . .                                     | 164 |
| 5.3 | Cohort 1 sepsis patient sepsis source . . . . .                                    | 165 |
| 6.1 | Multi-omic integration methods . . . . .                                           | 210 |
| A.1 | BioLegend TotalSeq-C CITE-seq antibodies. . . . .                                  | 229 |
| A.2 | BD Rhapsody Abseq antibodies. . . . .                                              | 231 |
| B.1 | Top differentially expressed genes between SRS . . . . .                           | 234 |
| B.2 | Top differentially expressed genes between sepsis-HV . . . . .                     | 235 |
| B.3 | Top differentially expressed genes between all-cause vs. COVID-19 sepsis . . . . . | 236 |
| B.4 | Top differentially expressed genes between critical vs. mild COVID-19 . . . . .    | 237 |
| C.1 | WGCNA B cell module genes . . . . .                                                | 240 |
| C.2 | WGCNA CD4+ T cell module genes . . . . .                                           | 241 |
| C.3 | WGCNA CD8+ T cell module genes . . . . .                                           | 242 |
| C.4 | WGCNA cMono module genes . . . . .                                                 | 243 |
| C.5 | WGCNA ncMono cell module genes . . . . .                                           | 244 |
| C.6 | WGCNA NK cell module genes . . . . .                                               | 246 |
| C.7 | WGCNA PB module genes . . . . .                                                    | 246 |
| C.8 | WGCNA module annotations . . . . .                                                 | 248 |
| D.1 | cNMF mature neutrophil programs (1-7) . . . . .                                    | 254 |
| D.2 | cNMF mature neutrophil programs (8-14) . . . . .                                   | 256 |
| D.3 | cNMF <i>S100A8/9</i> high neutrophil programs . . . . .                            | 258 |
| D.4 | cNMF <i>IL1R2</i> + immature neutrophil programs . . . . .                         | 260 |
| D.5 | cNMF <i>PADI4</i> + immature neutrophil programs . . . . .                         | 262 |
| D.6 | cNMF <i>MPO</i> + immature neutrophil/progenitor programs (1-6) . .                | 264 |
| D.7 | cNMF <i>MPO</i> + immature neutrophil/progenitor programs (7-12) . .               | 266 |

## ABBREVIATIONS

---

|                       |                                                                |
|-----------------------|----------------------------------------------------------------|
| <b>Ab</b>             | Antibody                                                       |
| <b>Abseq</b>          | BD Rhapsody cell surface protein sequencing                    |
| <b>ADT</b>            | Antibody-derived tags                                          |
| <b>APC</b>            | Antigen presenting cell                                        |
| <b>ARDS</b>           | Acute respiratory distress syndrome                            |
| <b>bp</b>             | Base pairs                                                     |
| <b>CAP</b>            | Community acquired pneumonia                                   |
| <b>cDC</b>            | Conventional dendritic cell                                    |
| <b>cDNA</b>           | Complementary deoxyribonucleic acid                            |
| <b>CI</b>             | Confidence interval                                            |
| <b>cMono</b>          | Classical monocyte                                             |
| <b>CITE-seq</b>       | Cellular indexing of transcriptomes and epitopes by sequencing |
| <b>COMBAT</b>         | COVID-19 multi-omic blood atlas                                |
| <b>cNMF</b>           | Consensus non-negative matrix factorization                    |
| <b>COVID/COVID-19</b> | Coronavirus disease 2019                                       |
| <b>CRP</b>            | C-reactive protein                                             |
| <b>CytoF</b>          | Cytometry by time of flight (mass cytometry)                   |
| <b>DAMP</b>           | Damage associated molecular pattern                            |
| <b>DA</b>             | Differential abundance                                         |
| <b>DC</b>             | Dendritic cell                                                 |
| <b>DE</b>             | Differential expression                                        |
| <b>DEG</b>            | Differentially expressed genes                                 |
| <b>DGE</b>            | Differential gene expression                                   |
| <b>DNA</b>            | Deoxyribonucleic acid                                          |
| <b>EDTA</b>           | Ethylenediaminetetraacetic acid                                |
| <b>eQTL</b>           | Expression quantitative trait loci                             |
| <b>FP</b>             | Fecal peritonitis                                              |
| <b>GEX</b>            | Gene expression                                                |
| <b>GO</b>             | Gene Ontology                                                  |
| <b>GOBP</b>           | Gene Ontology biological processes                             |
| <b>GWAS</b>           | Genome wide association study                                  |
| <b>FDR</b>            | False discovery rate                                           |
| <b>FC</b>             | Fold change                                                    |
| <b>FACS</b>           | Fluorescence activated cell sorting                            |
| <b>HLA</b>            | Human leukocyte antigen                                        |
| <b>HSPC</b>           | Hematopoietic stem and/or progenitor cell                      |
| <b>ICU</b>            | Intensive care unit                                            |
| <b>IL</b>             | Interleukin                                                    |
| <b>IRF</b>            | Interferon regulatory factor                                   |
| <b>LFC/LogFC</b>      | Log fold change                                                |
| <b>MALT</b>           | Mucosal associated lymphoid tissue                             |
| <b>MDSC</b>           | Myeloid derived suppressor cell                                |
| <b>MHC</b>            | Major histocompatibility complex                               |
| <b>MNP</b>            | Mononuclear phagocyte                                          |

## LIST OF TABLES

---

|                   |                                                 |
|-------------------|-------------------------------------------------|
| <b>ncMono</b>     | Non-classical monocyte                          |
| <b>NET</b>        | Neutrophil extracellular trap                   |
| <b>NEWS2</b>      | National early warning score 2                  |
| <b>Nhood</b>      | Neighborhood                                    |
| <b>NK cell</b>    | Natural killer cell                             |
| <b>PBMC</b>       | Peripheral blood mononuclear cell               |
| <b>PAMP</b>       | Pathogen associated molecular pattern           |
| <b>PC</b>         | Principal component                             |
| <b>PCA</b>        | Principal component analysis                    |
| <b>PMA</b>        | Phorbol 12-myristate 13-acetate                 |
| <b>PCR</b>        | Polymerase chain reaction                       |
| <b>pDC</b>        | Plasmacytoid dendritic cell                     |
| <b>qSOFA</b>      | Quick sequential organ failure assessment       |
| <b>QC</b>         | Quality control                                 |
| <b>RNA</b>        | Ribonucleic acid                                |
| <b>RNA-seq</b>    | RNA-sequencing                                  |
| <b>SARS-CoV-2</b> | Severe acute respiratory syndrome coronavirus 2 |
| <b>scRNA-seq</b>  | Single-cell RNA-sequencing                      |
| <b>SI</b>         | The Sepsis Immunomics Study                     |
| <b>SIRS</b>       | Systemic Inflammatory Response Syndrome         |
| <b>SLO</b>        | Secondary lymphoid organ                        |
| <b>SOFA</b>       | Sequential organ failure assessment             |
| <b>SRS</b>        | Sepsis response signature                       |
| <b>TF</b>         | Transcription factor                            |
| <b>TLR</b>        | Toll-like receptor                              |
| <b>WGS</b>        | Whole genome sequencing                         |
| <b>WGCNA</b>      | Weighted gene correlation network analysis      |
| <b>WHO</b>        | World Health Organization                       |
| <b>WHG</b>        | Wellcome Centre for Human Genetics              |

## INTRODUCTION

*This chapter presents the aims of this thesis, and provides an overview of pre-existing knowledge relevant to these goals*

|     |                                                                                                             |    |
|-----|-------------------------------------------------------------------------------------------------------------|----|
| 1.1 | The clinical syndrome of sepsis . . . . .                                                                   | 2  |
| 1.2 | SARS-CoV-2 and COVID-19 . . . . .                                                                           | 8  |
| 1.3 | The immune system . . . . .                                                                                 | 12 |
| 1.4 | Prospective cohort studies – the GAINs and Sepsis Immunomics<br>Studies and the COMBAT consortium . . . . . | 18 |
| 1.5 | Functional genomics in sepsis and COVID-19 . . . . .                                                        | 21 |
| 1.6 | Specific aims and objectives . . . . .                                                                      | 30 |

The overall aim of this thesis is to investigate the pathophysiology of sepsis using an integrated immunological and -omic approach at single cell resolution. In particular, this thesis examines sepsis from both an overall perspective of the condition, and in the context of a specific etiology seen in coronavirus disease 2019 (COVID-19).

Sepsis is a clinical syndrome defined as the dysregulated host response to infection leading to life threatening organ dysfunction (Singer et al. 2016). Substantial inter-individual variation and limited understanding of the specific disease processes and pathophysiological mechanisms behind the syndrome have led to a lack of immunomodulatory therapies until COVID-19. However, despite successes in COVID-19, the issues of heterogeneity remains unresolved in terms of all-cause sepsis (Marshall 2014).

In this thesis, I investigate the molecular and cellular immune features of the

host response in sepsis by integrating data from multiple -omic approaches at both bulk and single cell resolution. I focus mostly on transcriptomics, but also analyze and present data at the cell surface protein level. The patient cohorts discussed in the thesis were derived from the Sepsis Immunomics Study which I set up as part of the thesis.

In this chapter, I provide the background context for the thesis in terms of understanding sepsis, COVID-19, the immune system, the relevant patient cohorts and a brief overview of the -omic techniques employed throughout.

## 1.1 The clinical syndrome of sepsis

### 1.1.1 Definitions of sepsis

The initial 1991 consensus for sepsis defined the condition as a systemic inflammatory response due to an infection (Bone et al. 1992), where the four original Systemic Inflammatory Response Syndrome (SIRS) criteria were:

1. tachycardia ( $>90$  beats/min)
2. tachypnoea (respiratory rate  $>20$  breaths/min)
3. fever or hypothermia (temperature  $>38$  or  $<36^{\circ}\text{C}$ )
4. leukocytosis, leukopenia, or bandemia (white blood cells  $>1,200/\text{mm}^3$ ,  $<4,000/\text{mm}^3$  or bandemia  $\geq 10\%$ )

The diagnosis of sepsis required two or more Systemic Inflammatory Response Syndrome (SIRS) criteria in addition to confirmed or suspected infection. In this original definition of sepsis (Sepsis-1) and in the subsequent Sepsis-2 (Levy et al. 2003), the term severe sepsis denoted sepsis with organ dysfunction, and septic shock as "sepsis-induced hypotension persisting despite adequate fluid resuscitation".

Since 2016, the new definitions from Sepsis-3 have superseded the original sepsis criteria. Sepsis-3 instead uses the sequential organ failure assessment (SOFA) and quick SOFA (qSOFA) scores to determine sepsis, replacing SIRS criteria (Singer et al. 2016). By doing so, the new definitions effectively eliminated the category of sepsis without organ dysfunction.

Sepsis is defined currently according to Sepsis-3 criteria as "life-threatening organ dysfunction caused by a dysregulated host response to infection". Septic shock is reserved for patients with circulatory, cellular and metabolic abnormalities associated with a higher mortality, which is defined clinically as requiring a vasopressor to maintain a mean arterial pressure of  $\geq 65$  mmHg and serum lactate  $\geq 2$  mmol/L in the absence of hypovolemia. With these clinical features, mortality rate is in excess of 40%.

### **1.1.2 Morbidity and mortality of sepsis**

As the most common cause for medical admissions into intensive care units (ICU), mortality of sepsis is high. A third of patients admitted to ICU die from sepsis within 30 days, and overall in-hospital mortality is currently estimated at 20-30% (Singer et al. 2016; Bauer et al. 2020; Fleischmann-Struzek et al. 2020).

In a recent report from the Global Burden of Disease Study, it was highlighted that the global burden of sepsis is substantial, with over 50 million people worldwide affected every year. With an estimated mortality of 11 million in 2017, this represented approximately one fifth of deaths worldwide (Rudd et al. 2020). A key risk factor for mortality is age, where hospital mortality in patients over 80 years of age is double that of young patients under 50 (Kotfis et al. 2019).

Moreover, there is a substantial morbidity and mortality for survivors of sepsis. In addition to around one sixth of sepsis survivors not surviving past the first year post sepsis, longer term life expectancy is also reduced: compared to age-matched populations without sepsis, recovered patients have increased mortality

at 2 and 5 years (Prescott et al. 2016; Rahmel et al. 2020). Close to 40% of the patients are also re-admitted to hospital within the first twelve months at least once (Shankar-Hari et al. 2020) with a high risk of re-infection (Wang et al. 2014). Aside from mortality and hospitalizations, there is increased nursing care dependency in sepsis survivors (Fleischmann-Struzek et al. 2021) and reduced quality of life (Cuthbertson et al. 2013).

### **1.1.3 Overview of sepsis pathophysiology**

A range of immune and non-immune organ systems may be disrupted in sepsis. Immune dysfunction is characterized by the typically simultaneous occurrence of hyperinflammatory features (especially elevated cytokine levels) and immunosuppression in both innate and adaptive arms of immunity, more of which is discussed below as part of the introduction to the immune system and in the subsequent chapters (Poll et al. 2021).

In addition to immune disruption, it has been posited that endothelial barrier dysfunction is another key axis of the dysregulation of the host response, especially in severe COVID-19 / SARS-CoV-2 viral sepsis (Ince et al. 2016; Boissrame-Helms et al. 2013; Hotchkiss et al. 2016; Opal and Poll 2015; Huertas et al. 2020). Disruption of the continuous endothelial layer normally coating the vasculature as well as the glycocalyx (glycoprotein-polysaccharide layer covering the endothelium) leads to release of inflammatory mediators which can result in immune activation as well as kidney, respiratory and liver injury. Meanwhile, the loss of barrier integrity leads to intravascular protein and plasma fluid loss to extravascular spaces, leading to oedema and distributive shock.

Another hypothesis for sepsis pathogenesis is that there is general metabolic shutdown (Singer et al. 2004; Singer 2008; Singer 2017; Stanzani et al. 2020). This is motivated by the apparent lack of cell death in sepsis despite organ dysfunction, leading Singer and colleagues to propose that there may be a state



of hibernation which should theoretically be reversible after resolution of the immune insult. In support of this, it has been observed that patients who succumb to sepsis have lower levels of ATP in muscle tissue (Brealey et al. 2002).

Indeed, there are other metabolic disturbances in sepsis that exist regardless of the validity of the shutdown hypothesis. There is evidence for a general shift in multiple tissues towards anaerobic glycolysis, as well as insulin resistance and elevated proteolysis and lipolysis (Andreis and Singer 2016). Sepsis monocytes have been shown to have both impaired glycolysis and oxidative phosphorylation as well (Cheng et al. 2016).

An alternative view which combines the relevance of immunity and bioenergetics was presented by Deutschmann and Tracey (Deutschman and Tracey 2014) where the authors suggest that sepsis represents a general failure of homeostasis, particularly in neuroendocrine and immune systems. One of the prime examples that combines both is the "inflammatory reflex" which modulates TNF release after LPS injection models (Tracey 2002; Pavlov and Tracey 2012; Pavlov et al. 2018). Signals such as LPS, TNF, IL-1 and other inflammatory cytokines are detected by sensory vagus nerve neurons which send afferent signals to the nucleus tractus solitarius in the brainstem. Efferent signals then travel back through the vagus nerve to the celiac ganglion which connects to adrenergic cells making up the splenic nerve. This in turn communicates with T cells which release acetylcholine and inhibits NF- $\kappa$ B signaling in marginal zone macrophages, thereby reducing TNF/IL-1 $\beta$  release from these cells (Rosas-Ballina et al. 2008). In animal models e.g. with rats, this reduces the lethality of shock induced by LPS injection (Borovikova et al. 2000). Clinically, the neuroimmune links are seen with an association of higher mortality in patients admitted with shock and signs of infection with reduced vagal nerve activity (Barnaby et al. 2002).

Neuroendocrinologically, there are disruptions to thyroid hormone signaling

leading clinically to euthyroid sick syndrome (Luo et al. 2017), as well as impaired glucocorticoid responses which are normally important for reducing cytokine release (Deutschman and Tracey 2014). Ghrelin, the "hunger" hormone often associated with food intake, has also been implicated in murine models of sepsis where cecal ligation and puncture (CLP) reduces serum ghrelin levels, while exogenous supplementation of ghrelin attenuates cytokine release via increased vagal nerve signaling (Wu et al. 2007).

Ultimately, the disruptions to the above systems result in downstream organ dysfunction (Hotchkiss et al. 2016). This can manifest in the form of hemodynamic instability with contributions from distributive shock and myocardial depression, which may be due to the toxic effects of cytokines on cardiomyocytes (Sato and Nasu 2015). Kidney injury is another frequent feature of sepsis although the exact mechanisms linking sepsis to acute tubular necrosis are poorly described (Ricci et al. 2011). Neurologically, patients often present with altered mental status or even delirium (Atterton et al. 2020). Even when the source of sepsis is not pulmonary, patients frequently experience tachypnea and may suffer from acute respiratory distress syndrome (Kim and Hong 2016).

### **1.1.4 Heterogeneity in sepsis**

It is increasingly accepted that there is substantial individual variation in the sepsis response which is not captured by clinical disease definitions or bedside assessment (Stanski and Wong 2019; Leligdowicz and Matthay 2019). This heterogeneity limits therapeutic options and is a major contributory factor in the failure of more than 100 randomised clinical trials of agents aiming to modulate the host response in sepsis (Marshall 2008; Marshall 2014). Within the context of the COVID-19 pandemic there has been significantly more success, with both corticosteroids and IL-6 receptor inhibitors having been shown to have positive effects on patient outcome (Abani et al. 2021). This is likely due to

reduced heterogeneity compared to sepsis as a single umbrella syndrome given the specific pathogen, and also careful patient stratification - for example, the use of corticosteroids is only recommended for patients requiring supplemental oxygen and can even cause harm in those without such a need, and IL-6 inhibition was selectively tested only in those with hypoxia i.e. requiring oxygen therapy or with signs of systemic inflammation as defined by C-reactive protein (CRP) levels  $\geq 75$  mg/L.

However, in the setting of sepsis overall, there are still no immunomodulatory therapies. It is likely that the nature of the immune response state at any given point during a patient's septic episode must be correctly appraised to address the issue of heterogeneity and to enable such drugs to be deployed effectively (Salama et al. 2021; Abani et al. 2021).

Multiple factors intersect to cause the heterogeneity seen in sepsis. A seminal twin study showed that the risk of dying from an infection was higher when a biologic parent died from infection, but not with an adoptive parent, suggesting that host genetics are key to infectious disease susceptibility (Sørensen et al. 1988). More broadly, the host genetic background affects infection susceptibility through a wide range of mechanisms, from single gene defects with high effect sizes as with Toll-like receptor (TLR) or interferon regulatory factor (IRF) mutations, which can lead to severe infection and sepsis (Casanova and Abel 2021), to more polygenic mechanisms (Kwok et al. 2021) e.g. as demonstrated in eQTL mapping for sepsis patient subgroups (Davenport et al. 2016). Host epigenetics, existence of comorbidities, and social and demographic factors may also interact with other environmental factors such as therapy given, where both timing and suitability of the treatment play a role. Microbial factors including the causative organism, virulence of the microbe and any antimicrobial resistance of the pathogen are also important (Goh and Knight 2017). For example, fungal infections such as invasive Candidiasis are notorious for particularly high

mortality rates (Delaloye and Calandra 2014).

### **1.1.5 Precision medicine approaches for sepsis**

Precision medicine refers to approaches in diagnostics, prognostics and therapeutics which account for the specific molecular features that define a patient's particular illness (Shankar-Hari and Rubenfeld 2019). Two specific aspects to the successful delivery of precision medicine include prognostic and predictive enrichment, i.e. selecting patients who are either more likely to die in the former, or are more likely to respond to therapy in the latter (Poll et al. 2021). In the case of predictive enrichment, this is usually due to therapy targeting a specified biological mechanism.

Given the heterogeneity in sepsis and relatively poor understanding of disease processes underlying sepsis, it is unsurprising that there is currently a lack of precision medicine approaches for the condition. The stratification of sepsis patients into biologically and clinically meaningful subgroups is one of the key questions that will be tackled in this thesis, in particular in chapters three and five.

## **1.2 SARS-CoV-2 and COVID-19**

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to a global pandemic of coronavirus disease 2019 (COVID-19) since late 2019/early 2020 with devastating impact on lives and livelihoods. As of time of writing, the world has seen over 6 million deaths with World Health Organisation (WHO) estimates that 15 million lives have been lost due to COVID-19. Despite the deployment of multiple vaccines, there continues to be a large number of cases and deaths, especially in resource-poor countries.

### 1.2.1 Clinical features of COVID-19

SARS-CoV-2 and the resulting disease of COVID-19 presents with mostly respiratory symptoms, but can also affect other organ systems of the body, including the gastrointestinal tract, an association with increased thrombotic risk, and reports of neurological manifestations (Osuchowski et al. 2021). The most severe cases of COVID-19 are characterised by respiratory insufficiency requiring invasive mechanical ventilation.

This pandemic has further exemplified the issue of sepsis as COVID-19 in its severest form can manifest as sepsis. The current World Health Organisation (WHO) classifications of COVID-19 severity stratify the disease into:

1. Asymptomatic or Presymptomatic Infection: Individuals who test positive for SARS-CoV-2 using a virologic test (i.e., a nucleic acid amplification test [NAAT] or an antigen test) but who have no symptoms that are consistent with COVID-19.
2. Mild Illness: Patients with any of the various signs and symptoms of COVID-19 (e.g. fever, cough, sore throat, malaise, headache, muscle pain, nausea, vomiting, diarrhea, loss of taste and smell) but who do not have shortness of breath, dyspnea, or abnormal chest imaging.
3. Moderate Illness: Patients with show evidence of lower respiratory disease during clinical assessment or imaging and who have an oxygen saturation ( $\text{SpO}_2$ )  $\geq 94\%$  on room air at sea level.
4. Severe Illness: Patients with  $\text{SpO}_2 < 94\%$  on room air at sea level, a ratio of arterial partial pressure of oxygen to fraction of inspired oxygen ( $\text{PaO}_2/\text{FiO}_2$ )  $< 300$  mm Hg, a respiratory rate  $> 30$  breaths/min, or lung infiltrates  $\geq 50\%$ .
5. Critical Illness: Characterised by respiratory failure, septic shock, and/or

multiple organ dysfunction.

Thus, in the more severe forms of COVID-19, the patients suffer from SARS-CoV-2 viral sepsis. Indeed, a meta-analysis of COVID-19 as a cause of sepsis found that more than three-quarters of patients admitted to ICUs and over a third of general ward patients fulfill Sepsis-3 criteria (Karakike et al. 2021).

### 1.2.2 COVID-19 pathophysiology

COVID-19 is caused by SARS-CoV-2, a positive sense single-stranded RNA virus which belongs to the coronavirus family (Zhu et al. 2020). Seven coronaviruses are known to be pathogenic humans with four leading to common colds, while SARS-CoV resulted in the 2003 SARS outbreak mostly in China, and Middle East Respiratory Syndrome (MERS) CoV the 2012 MERS outbreak originating from Saudi Arabia (Su et al. 2016).

SARS-CoV-2 enters host cells chiefly through viral spike (S) protein - host cell angiotensin converting enzyme 2 (ACE2) interaction after transmembrane protease serine 2 (TMPRSS2) priming of the S protein (Hoffmann et al. 2020). The high expression of both ACE2 and TMPRSS2 in the nasal epithelium likely accounts for the high transmissibility of the virus (Sungnak et al. 2020; Hou et al. 2020). Moreover, the widespread expression of these host factors throughout the body (Hikmet et al. 2020; Muus et al. 2021) explains the wide range of symptoms that can afflict individuals infected with SARS-CoV-2 as the virus is able to infect many different tissues and organs including cardiocytes, endothelial, testicular and bile duct cells. Moreover, the virus is able to enter host cells beyond interacting with cell membrane ACE2 as soluble ACE2 and vasopressin can bind to virions which then leads to cellular uptake by receptor mediated endocytosis (Yeung et al. 2021).

The primary site of pathology in COVID-19 is the respiratory system. Pathological examination of patients who succumb to COVID-19 show edema

and congestion of the lungs macroscopically and type II pneumocyte damage microscopically (Carsana et al. 2020). Platelet-fibrin microthrombi can be seen in alveolar capillaries (Carsana et al. 2020), and there is frequent infiltration of leukocytes, with lymphocytes found in both interstitial and perivascular locations while monocytes are also found in the interstitium and macrophages within alveoli.

COVID-19 therefore unsurprisingly can lead to acute respiratory distress syndrome, defined as non-cardiogenic pulmonary edema and diffuse lung inflammation, estimated to occur in 50% of patients with severe disease (Osuchowski et al. 2021). Even after recovery, patients may be left with long-term lung fibrosis.

COVID-19 is also characterised by endothelial damage and coagulopathy. Estimates of venous thromboembolism (VTE) rates range between 21-69% (Obi et al. 2021), which is substantially higher than ICU patients post surgery (Obi et al. 2015). Clinical observational studies have reported features of hypercoagulability in COVID-19 patients including elevated plasma fibrinogen and D-dimer levels (Guan et al. 2020; Lippi and Plebani 2020). Endothelitis observed in COVID-19 has been hypothesised to contribute to refractory ARDS seen in COVID-19 (Ackermann et al. 2020); moreover, the endothelial dysfunction, which is a shared feature with sepsis, may contribute to the hypercoagulability in COVID-19 (Ackermann et al. 2020).

The host immune response in COVID-19 is also relevant to disease pathogenesis, with classical features of an immune response to infection seen (detailed more below). One notable feature that has received significant attention is the notion of a "cytokine storm" (Chen and Quach 2021), as reflected by elevated plasma cytokine levels which was detailed in the first report of clinical and laboratory features of COVID-19 patients from Wuhan, China (Huang et al. 2020), and subsequently also found in other patient cohorts. In the original report, patients

with COVID-19 were found to have elevated plasma IFN $\gamma$ , IL-1 $\beta$ , IP10 and MCP1 and ICU patients further had higher GCSF, IP10, MCP1, MIP1A, and TNF than non-ICU patients (Huang et al. 2020). IL-6, a pro-inflammatory cytokine, was subsequently also found to be elevated in COVID-19 patients, and represented an attractive target given the already clinically licensed use of drugs such as Tocilizumab which target the IL-6 receptor for various inflammatory conditions including rheumatoid arthritis, giant cell arteritis, haemophagocytic syndrome, or cytokine release syndrome (e.g. in the case of CAR-T cell therapies).

However, the concept of COVID-19 inducing a "cytokine storm" has not been without controversy, as comparisons of plasma TNF, IL-8 and IL-6 were all lower than levels seen in sepsis or other cohorts of ARDS patients (Sinha et al. 2020; Kox et al. 2020). Nevertheless, IL-6 pathway inhibition via IL6R inhibition by Tocilizumab has been shown in the RECOVERY trial to reduce the composite endpoint of invasive mechanical ventilation or death in patients who did not initially receive ventilatory support (Abani et al. 2021).

Indeed, suppressing exuberant inflammation in COVID-19 has been shown to be beneficial in patients in addition to IL-6 inhibition; dexamethasone glucocorticoid treatment was the first approved drug to show mortality benefits in COVID-19 patients requiring oxygen (The RECOVERY Collaborative Group 2021). Since the publication of positive results for dexamethasone and Tocilizumab in treating COVID-19 patients, the RECOVERY trial has also reported that baricitinib, a janus kinase 1/2 inhibitor, also reduces mortality in hospitalized COVID-19 patients (The RECOVERY Collaborative Group 2022).

### **1.3 The immune system**

The immune system is the body's defense against injury and infection (Murphy 2017). Broadly divided into the innate and adaptive arms, the system exists



as a collection of molecules and cells spread throughout the body, with some organs contributing a particularly important role such as the spleen and thymus. Primary lymphoid organs including the bone marrow and thymus are organs where immune cells are produced and mature, while secondary lymphoid organs (SLOs) i.e. lymph nodes, the spleen as well as various mucosal associated lymphatic tissues (MALTs) are responsible for initiation of adaptive immune responses.

Immune system dysfunction can lead to a range of infectious diseases or autoimmunity and allergies as well as development of malignancies.

### **1.3.1 The innate immune system**

The innate immune system at its most basic includes barrier defence mechanisms including epithelial cell lining and molecules such as defensins. The major immune cell type is the granulocyte, of which the most abundant in peripheral blood are neutrophils, as well as monocytes, macrophages, natural killer (NK) cells and dendritic cells (DCs).

Neutrophils and monocytes are responsible for phagocytosis of pathogens, while NK cells mediate antibody dependent cellular cytotoxicity (ADCC) and direct cell killing with granzymes and perforin, much like CD8<sup>+</sup> T cells. DCs meanwhile serve as the bridge between innate and adaptive immune systems, acting as professional antigen presenting cells (APCs) which activate naïve T cells in SLOs via major histocompatibility complex (MHC) antigen presentation and costimulatory molecule (CD80 and CD86) expression.

### **1.3.2 The adaptive immune system**

Whereas the innate immune system is responsible for early response and is pre-programmed regardless of the insult or pathogen, the adaptive immune system requires multiple days to be activated, but confers exquisite antigen specificity

and immune memory.

The key adaptive immune cells include T and B lymphocytes. These cells are characterised by their ability to undergo somatic recombination of their receptor sequences to theoretically produce up to  $10^{13}$  and  $10^{18}$  distinct repertoires for B and T cells respectively (Calis and Rosenberg 2014). Whereas B cells recognize conformational epitopes, T cells require peptide antigen to be presented on MHC molecules for T cell receptor (TCR) recognition. While exceptions exist, the overarching paradigm remains that CD4<sup>+</sup> T cells recognize extracellular peptide antigens via MHC class II presentation while CD8<sup>+</sup> T cells recognize intracellular peptide antigens presented on MHC class I molecules. To ensure that the potent adaptive immune system is not unleashed against the host, a wide range of central and peripheral tolerance mechanisms exist during lymphocyte development to select receptor repertoires that theoretically should not recognize self antigens.

A number of cells which resemble T lymphocytes in phenotype and function but do not have the same antigen receptor recombination as conventional T cells also exist and are termed either innate-like or unconventional T cells, including mucosal associated invariant T (MAIT) cells, NKT cells, invariant NKT (iNKT) cells, T cells restricted to CD1a/b/d, and T cells with  $\gamma\delta$  TCRs instead of  $\alpha\beta$  TCRs (Godfrey et al. 2015). These T cells recognize small-molecule metabolites, lipids and other modified peptides, and provide rapid, local effector responses which complement the sterilizing immunity that classic  $\alpha\beta$  T cells can deliver (Mayassi et al. 2021).

### 1.3.3 Dynamics of the immune response

Injury from infectious or non-infectious insults results in the process of inflammation which is the first reaction in our immune response. Acute inflammation is characterised by local redness and heat, swelling and pain.

Pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs), nucleotide-binding oligomerization domain and leucine-rich repeat containing gene family receptors (NLRs), C-type lectin receptors (CLRs), retinoic acid-inducible gene (RIG)-I-like receptors (RLRs) and other DNA sensing molecules (e.g. cyclic GMP–AMP synthase, cGAS) exist on resident immune cell surfaces or intracellularly (Takeuchi and Akira 2010; Li and Wu 2021). They detect pathogen or damage associated molecular patterns (P/DAMPs) to begin the process of acute inflammation. Signal transduction leads to activation of key pathways such as the nuclear factor  $\kappa$ B (NF- $\kappa$ B) pathway, which leads to soluble mediator release, including cytokines, chemokines and other vasoactive mediators such as histamine, bradykinins and prostanoids (Chen et al. 2018; Ricciotti and Fitzgerald 2011). This in turn leads to vasodilatation, accumulation of inflammatory exudate, and recruitment of more immune cells. The complement cascade is also activated to assist in destroying pathogens (through opsonization and membrane attack complex formation) and in recruiting immune cells (through anaphylatoxin activity of C3a, C4a and C5a) (Reis et al. 2019). A particular class of cytokines is the interferons (IFNs), which are split into three types, of which type I and III IFNs are classically associated with antiviral functions, while class II consists of a single cytokine IFN $\gamma$  which is important in directing cellular immunity by stimulating NK cells, macrophages as well as neutrophils (McNab et al. 2015; Lee and Ashkar 2018).

As leukocytes arrive, they begin the processes of removal of the injurious agent. If the pathogen encountered cannot be handled by the innate immune system alone, an adaptive immune response is deployed when DCs present antigen in SLOs to select cognate T and B cells clonally (Adams et al. 2020). Clonal expansion of the relevant adaptive immune cells brings in humoral immunity through B cell/plasma cell antibody production and cellular immunity through CD8+ T cell cytotoxicity. Antibodies produced by B cells can neutralize antigens

e.g. viral particles, opsonize antigens for enhanced phagocytosis, and mediate ADCC (Lu et al. 2017). CD8+ T cells kill target cells after recognising peptide-major histocompatibility complex (MHC) complexes by releasing molecules such as perforin and granzymes. CD4+ T cells secrete cytokines and coordinate the adaptive immune response by providing relevant "help" to CD8+ T cells and B cells; e.g. licensing the former for enhanced cytotoxicity and memory formation (Sun and Bevan 2003; Smith et al. 2004), and helping in germinal center reactions for the latter for B cell affinity maturation and antibody class switching (Crotty 2015).

#### **1.3.4 Proinflammatory mechanisms vs. immunosuppression in sepsis**

The immunological responses in sepsis have been characterized thoroughly in terms of cellular and molecular mechanisms. While the response largely has similar molecular and cellular features to those of infections in general, there is debate regarding the imbalances in proinflammatory mechanisms vs. immunosuppression (Poll et al. 2021), and the contributions of either to pathology and phenotype. This is reflected in the notion of dysregulation in host response as defined currently under the Sepsis-3 consensus.

Early animal model studies of intravenous bacterial product challenge, often lipopolysaccharide (LPS), showed initial release of proinflammatory cytokines including TNF and IL-1 $\beta$  (Wiersinga et al. 2013). Although not an ideal mimic of the human condition of sepsis, the release of these cytokines is likely to play a role in sepsis pathology. Other proinflammatory soluble mediators relevant to sepsis include the complement and coagulation systems which together result in "immunothrombosis" (Engelmann and Massberg 2013), where microvasculature thromboses can eventually lead to the clinical condition of disseminated intravascular coagulation (DIC). Aside from organ failure due to

thrombosis and poor blood supply to the affected organ, the consumption of coagulation factors can also lead to simultaneous bleeding in other areas of the body.

Cellular components also have proinflammatory contributions to sepsis pathogenesis. In the peripheral blood of sepsis patients, neutrophilia is observed. One effector function of neutrophils is to release neutrophil extracellular traps (NETs) to ensnare and kill pathogens (Papayannopoulos 2017). These NETs comprise a mix of chromatin and proteases such as cathepsin G and elastase which give the traps their antimicrobial activity. However, NETosis can also lead to thrombosis and vascular damage due to the histones from the chromatin that form a major part of NETs (Denning et al. 2019).

Conversely, there are mechanisms of immunosuppression at play in sepsis. Clinically, this manifests as secondary infection risk, occasionally by opportunistic pathogens, and even reactivation of viruses (Walton et al. 2014). In addition to the neutrophilia in peripheral blood, there is lymphopenia which correlates with disease severity and mortality risk (Drewry et al. 2014; Vahedi et al. 2019; Cilloniz et al. 2021), and there is evidence that T cells from sepsis patients have reduced ex vivo effector function capacities including cytokine secretion (IL-2 and IFN $\gamma$ ), proliferation and cytotoxic function in the case of CD8 $^{+}$  T cells (Boomer et al. 2011). Together with elevated PD-1 expression in these T cells, there has been suggestion of T cell exhaustion in sepsis. Studies of immune cells from post mortem examination of sepsis patients also identified that a range of mononuclear cell types including CD4 $^{+}$  and CD8 $^{+}$  T cells, NK cells and follicular dendritic cells all exhibit increased apoptosis (Hotchkiss et al. 1999)

Finally, in the mononuclear phagocyte compartment specifically, there is evidence for reduced HLA-DR (class II MHC) expression in sepsis at both gene and protein level, which would lead to reduced antigen presentation and

suppressed immune function and is indeed associated with increased risk of nosocomial infection and death (Monneret et al. 2006; Landelle et al. 2010). This has been related to the in vitro and in vivo experimental phenomenon of endotoxin tolerance, which describes the hyporesponsiveness to LPS challenge after an earlier low dose exposure to LPS (Beutler and Rietschel 2003). Importantly, this phenomenon can be "crossed" i.e. different pathogen or damage associated molecular patterns (PAMPs/DAMPs) can also lead to the reduced response to LPS (Aneja et al. 2008; Austermann et al. 2014). Enrichment of endotoxin tolerance gene expression signatures (Pena et al. 2011) seen in sepsis patients with poor clinical outcomes has previously been shown, suggesting a component of mononuclear cell dysfunction in sepsis (Pena et al. 2014; Davenport et al. 2016).

Together, these mononuclear cellular features suggest dampened immune function or activity in sepsis patients, leading to the hypothesis of immunosuppression in the condition. However, it is clear that the immune dysregulation is not a binary property of either excessive inflammation or immunosuppression, given that features of both exist simultaneously, and that data exists that does not necessarily fit with each hypothesis e.g. the low attributable mortality of secondary infections to sepsis (Van Vught et al. 2016).

## **1.4 Prospective cohort studies – the GAINs and Sepsis Immunomics Studies and the COMBAT consortium**

Established in 2005, the UK Genomic Advances in Sepsis (GAINs) study was set up to enable genetic and multi -omic studies in septic patient cohorts (Rautanen et al. 2015). The study has recruited over 2000 patients with sepsis due to community acquired pneumonia (CAP) or faecal peritonitis (FP)

with genotyping, peripheral blood total leukocyte gene expression and plasma proteomic data having been generated.

To build on the disease endotype work established by peripheral blood total leukocyte transcriptomic analyses in septic patients (Davenport et al. 2016; Burnham et al. 2017), the next step was to identify cellular drivers behind the signatures observed. In order to understand this, experiments requiring cellular sample material needed to be performed. However, the study ethics for GAINs precluded any cellular material to be isolated or sorted for downstream assays examining cell type specificities. There was therefore a need to establish a follow-on study.

### **Setup of Sepsis Immunomics Study**

A new application for Health Research Authority (HRA) and Health and Care Research Wales (HCRW) approval and research ethics committee (REC) review to establish such a study was prepared and submitted to the Integrated Research Application System (IRAS) (IRAS project ID 260007) on 15 May 2019. The study was titled “The GAINs Investigators: application of an integrated immune -omic approach in sepsis” or in short “Sepsis Immunomics”. REC approval was gained after a meeting with the South Central – Oxford C Research Ethics Committee on 28 June 2019, with Health Research Authority (HRA) and Health and Care Research Wales (HCRW) approval confirmed on 15 July 2019 (REC reference 19/SC/0296).

Study protocol design adhered strictly to the 2005 Mental Capacity Act for consenting procedures given that a large proportion of septic patients are mentally obtunded and do not have capacity to consent. Key features of the study protocol extended beyond the GAINs study protocol relevant to the current thesis included:

1. an Emergency Waiver of Consent under the 2005 Mental Capacity Act

Section 32(9) to allow for early blood sampling on arrival to the hospital at the emergency department before consent can be taken, as well as a similar waiver for the first in-hospital sample. Such a model was proposed for the study as the disease process in sepsis evolves rapidly and is time sensitive. To enable clinical decision making based on biology at the earliest time points, sampling during the initial stages is essential. In addition, sampling prior to administration of antimicrobial therapy is important to capture biological information obscured by treatment. However, the realities of clinical care mean that demanding for consent procedures during the early period of hospital admission, especially prior to antimicrobial therapy, would either impede clinical care or put an undue burden on healthcare staff and likely lead to insufficient recruitment.

2. consent for more complex sample types to be taken, including sampling and storage of cellular material such as whole blood total leukocytes, or peripheral blood mononuclear cell (PBMC) separation by density gradient centrifugation and cryopreservation. This change encompassed the scope to perform multi -omic experiments for which previous ethics approval was not in place, including single cell RNA-sequencing, immunophenotyping, and epigenetic experiments.
3. the addition of a sample up to six months after hospital discharge to enable evaluation of the long term, especially epigenetic aspects of sepsis on the immune system.
4. the extension of recruitment to patients beyond ICUs to include patients from the emergency department / general wards, as a large number of septic patients who are not admitted to ICU are nonetheless critically ill.

In early 2020, as the COVID-19 outbreaks began reaching the scale of a pandemic, the SI study ethics were amended to allow COVID-19 cases to be recruited under the same sampling scheme, with the exception that all



hospitalized COVID-19 patients could be recruited even without the diagnosis of sepsis.

### **The COMBAT consortium**

The COMBAT (COVID-19 Multi-Omic Blood ATlas) Consortium was an effort established in 2020 and led by the University of Oxford to perform deep immune phenotyping of peripheral blood of COVID-19 patients. The aims were to use multiple -omic modalities combined with detailed clinical phenotyping to understand the features and correlates of disease in the host response to SARS-CoV-2. The majority of patients were recruited under the Sepsis Immunomics study, with a small number of samples also from other cohorts within Oxford under the Translational Gastroenterology Unit Biobank or from collaborations in London under the Aspergillosis in patients with severe influenza (AspiFlu ISRCTN51287266) study [Wales REC 5, reference 19/WA/0310].

## **1.5 Functional genomics in sepsis and COVID-19**

Functional genomics is a branch of biology which aims to understand how regions of the genome lead to biological processes (Guigo and Hoon 2018). It operates at a genome wide level rather than a gene-by-gene approach as often employed in classical molecular biology.

The range of molecular information that comes under functional genomics is broad, from genetic to transcriptomic data, as well as epigenomic and proteomic data. The aim is therefore to understand biology at the level of systems.

### **1.5.1 Genome wide association studies in sepsis**

As DNA contains the instructions for an organism, sequence level information is the most fundamental level of information in understanding genome function. Outside of monogenic causes (Casanova and Abel 2021), susceptibility to

infectious disease can be considered as a complex trait (Kwok et al. 2021), with genetic contributions from an aggregation of low penetrance but common variants (i.e. the common variant-common disease hypothesis) (Schork et al. 2009). Genome-wide association studies (GWAS) have been instrumental in dissecting the genetic basis of common traits, uncovering genes related to phenotype by comparing the frequency of genetic variants in populations of different phenotypes, e.g. patients vs. healthy controls and highlighting the variants that are statistically associated with the traits. This has uncovered underlying biology of the phenotype, enabled genetic correlations to be calculated, and provided an opportunity for clinical risk prediction, most notably with polygenic risk scores (Tam et al. 2019; Uffelmann et al. 2021).

At the time of writing, there have been four GWAS conducted for sepsis. The first GWAS by Rautanen et al (2015) examined 28-day mortality in patients with community acquired pneumonia (CAP) sepsis. A single hit which reached genome-wide significance in the *FER* gene was associated with reduced risk of death (age-adjusted odds ratio 0.56, 95% CI 0.45-0.69). *FER* is a tyrosine kinase that can promote the activation of NF- $\kappa$ B, and is also important in leukocyte recruitment and diapedesis. However, neither a targeted validation (Schöneweck et al. 2015) or subsequent GWAS have been able to replicate this signal.

Two other GWAS then uncovered variants in *VPS13A* (role in autophagy) (Scherag et al. 2016) and *CISH* (role in cytokine responses) (Rosier et al. 2021) that were associated with increased 28-day mortality. By contrast, variants in *SAMD9* and two intergenic regions were associated with increased 28-day survival (Hernandez-Beeftink et al. 2022).

The number of genes associated with sepsis outcomes is therefore low in comparison to many other complex diseases. Moreover, as in GWAS of other diseases, variants are often in non-coding regions i.e. the functional impact is not easily interpretable. It is likely that the effect of these non-coding variants is

to influence the expression of genes (Maurano et al. 2012). The paucity of genes related to sepsis outcome is likely due to relatively low sample sizes in sepsis GWAS e.g. 2534 patients in the initial GWAS (Rautanen et al. 2015), compared to the latest GWAS in schizophrenia with 76,755 cases and 243,649 controls (Trubetskoy et al. 2022), significant heterogeneity of patients as discussed above, and a possibility that the genetic determinants of sepsis outcome may simply not be particularly strong.

### 1.5.2 Genetic susceptibility in COVID-19

By contrast, identification of genetic variants that affect COVID-19 outcomes have been more successful due to larger studies with sufficient sample size which has been enabled by biobanks with extensive genotyping data e.g. the UK biobank, 23andme (Shelton et al. 2021) and AncestryDNA (Roberts et al. 2022), but likely also at least in part due to the homogeneity of the condition i.e. there is a single pathogen leading to the disease process.

The first GWAS conducted (Group 2020) examined genetic variants associated with severe disease i.e. respiratory failure, discovering two loci including a signal at the ABO blood group locus and one at 3p21 situated near a gene cluster containing immune related cytokine genes e.g. *CCR9*, *CXCR6*, *XCR1* and also other genes (*SLC6A20*, *LZTFL1*, *FYCO1*). These signals were robustly reproduced in subsequent studies, and multiple other loci have been discovered or corroborated since, in particular by the COVID-19 Host Genetics Initiative (Niemi et al. 2021) which has been continuously adding new loci with each new data freeze. This includes a number of genes related to the interferon response including loci 12q24 (*OAS1*, *OAS2* and *OAS3*) and 21q22 (*IFNAR2*) which were first reported by Pairo-Castineira et al (2020). Perhaps unsurprisingly, variants linked to *ACE2* (Horowitz et al. 2022) and *TMPRSS2* (Roberts et al. 2022; Niemi et al. 2021) i.e. the main host factor involved in

COVID-19 entry into host cells have also been implicated. Interestingly, GWAS approaches also identified the locus of *UGT2A1/UGT2A2* as related to the loss of smell or taste reported in COVID-19 (Shelton et al. 2022). As the cost of sequencing continues to drop, whole genome sequencing (WGS) methods become more tractable; a recent attempt in COVID-19 revealed variants not directly implicated by GWAS but in pathways related to the genes uncovered through GWAS, such as *FUT2* i.e. blood-type antigen secretor status and *IL10RB* and *PLSCR1* which have roles in interferon signaling (Kousathanas et al. 2022).

In addition to the genome wide methods to interrogate genetic susceptibility, a more hypothesis and candidate gene-based approach has triangulated on the potential importance of the interferon response to COVID-19 outcomes. In a pair of studies, one looking at autoantibodies against type I IFNs and one utilising whole exome and whole genome sequencing for loci known to be important in inborn errors of immunity, Casanova and colleagues provided evidence for the importance of type I IFNs. First, it was identified that in approximately 10% of patients with life-threatening COVID-19, there was the presence in the blood of autoantibodies against either type I or type III IFNs (Bastard et al. 2020). In a handful of cases where plasma samples were available before infection, it was seen that the autoantibodies existed before the onset of infection. Simultaneously, it was demonstrated that patients with severe disease had an excess of nonsynonymous mutations with predicted loss of function in genes previously shown to lead to life-threatening disease from influenza (*TLR3* (Lim et al. 2019), *IRF7* (Ciancanelli et al. 2015), *IRF9* (Hernandez et al. 2018) and other genes that were in pathways connected to these hub genes (*TRIF*, *UNC93B1*, *TRAF3*, *TBK1*, *IRF3*, *NEMO*, *IFNAR1*, *IFNAR2*, *STAT1* and *STAT2* (Casanova and Abel 2021)), many of which were key in the type I IFN response (Zhang et al. 2020).

### 1.5.3 Transcriptomic profiling

In the first step of the central dogma of biology, DNA is read and messenger RNA (mRNA) produced in a process known as transcription, or gene expression. This process is tightly controlled both spatially and temporally. As proteins mediate many of the biological functions within cells, and protein translation is based on the mRNA molecules that are transcribed, we often approximate cellular function by examining the genes that are expressed in any given cell or tissue to infer biological function.

While interrogating the expression of specific genes with methods such as quantitative polymerase chain reaction (qPCR) can be useful if specific targets are suspected to be of interest, this has the drawbacks of potential bias, and difficulty in discovering novel aspects of biology systematically. The era of transcriptome profiling has seen a significant conceptual and technological revolution enabled by the Human Genome Project (HGP). Having the entire sequence of the human genome has given us the further ability to generate the whole transcriptome. The ability to interrogate the entire transcriptome has allowed gene expression profiling in an unbiased manner, leading to utility across various species, and both in basic biology and in more translational or clinical endeavours.

Bulk RNA-seq has become a mainstay in the toolkit for analysing gene expression due to much reduced sequencing costs in recent years (Stark et al. 2019). Compared to previous technologies interrogating the transcriptome based on hybridization methods i.e. microarrays, sequencing has two key advantages (Wang et al. 2009). Firstly, the sensitivity of detecting lowly expressed genes with sequencing is greater in RNA-seq than with microarrays. Secondly, RNA-seq allows all RNA molecules to be captured and therefore detection of de-novo transcripts, whereas microarrays are limited to capturing only genes which complementary probes have been designed for. RNA-seq has therefore become

the method of choice for gene expression analyses.

RNA-seq involves capturing RNA molecules either through a poly-A tail hybridization or ribosomal RNA deletion, the former capturing mRNA and the latter other species such as long non-coding RNAs or micro RNAs as well ("total" RNA-seq). The RNA is reverse transcribed into complementary DNA (cDNA), amplified into a sequencing library with adapters attached, and sequenced, often on Illumina sequencers. Reads are aligned to either a reference genome or transcriptome or pseudo-aligned, and read counts per transcript or gene are quantified.

#### **1.5.4 Analysis of transcriptomic data**

One key aim of transcriptomic experiments is to identify genes and corresponding pathways that are differentially expressed across different observational or perturbational conditions. In the setting where the condition is known, analysis can proceed with supervised methods e.g. differential gene expression (DGE) testing. However, transcriptomics may also be used to group patients in an unsupervised manner based on their gene expression patterns.

##### **Exploratory data analysis**

Exploratory data analysis (EDA) refers to methods aimed to understand the variation of data within samples, the correlation across samples and underlying statistical properties of the data e.g. in RNA-seq the relationship between mean and variance of count data. EDA is also synonymous with using unsupervised methods to query the dataset, with most notable examples being clustering methods and linear algebraic methods such as principal components analysis (PCA).

Clustering methods may be used to group patients in an unsupervised manner (Xu and Tian 2015). Commonly used methods in genomics and transcriptomics

include hierarchical clustering, k-means clustering or with more modern methods in terms of graph-based or spectral theory based methods. A shared feature of most clustering algorithms is need for a similarity/distance metric, of which euclidean distance is the most often used given a longstanding history and interpretability of the metric. Clustering techniques have previously identified disease subgroups in many different conditions ranging from systemic lupus erythematosus (Banchereau et al. 2016) to cancer, and indeed also in sepsis (Davenport et al. 2016).

PCA aims to decompose the variance structure of a dataset through matrix decomposition of the covariance matrix of the data. PCs that are factorized out are ranked in order of variance accounted for, and are orthogonal to each other. While strictly linear i.e. the method has a drawback of potentially missing non-linear relationships, the method is nevertheless useful given its speed, deterministic nature and the interpretability of linear methods. PCAs are useful for identifying subgroups of samples, occasionally when combined with clustering methods on PCs, and in assessing data quality, e.g. in identifying batch effects. A number of matrix decomposition techniques other than PCA exist for similar unsupervised analysis of genomic and transcriptomic data (Brunet et al. 2004; Stein-O'Brien et al. 2018).

### **Differential testing**

Following unsupervised analyses, DGE analysis is often performed across either pre-determined experimental groups or groups as identified from unsupervised EDA. In the context of observational studies as in this thesis, it is not possible to differentiate between whether the changes are causes or consequences of disease, but such analysis can nevertheless identify correlates of disease and therefore changes in biological pathways that may be relevant do pathogenesis.

A number of statistical frameworks have been developed to model RNA-seq

data for across sample comparisons. It is now widely accepted that negative binomial models are best suited for RNA-seq count data (Robinson et al. 2010; Love et al. 2014) or that a transformation is needed to account for overdispersion in the mean-variance relationship before direct linear modeling (Law et al. 2014), and that normalization requires consideration of both library size and RNA composition differences (Conesa et al. 2016). DGE analysis under these frameworks have been important in identifying pathways relevant to disease in infectious disease (Thair et al. 2021), neurodegeneration (Marques-Coelho et al. 2021) and cancer (Shaheen et al. 2018; Stewart et al. 2013).

### **Modular techniques**

Aside from testing for differences in expression of each individual gene in a transcriptomic dataset between different comparator groups, there have been a number of methods for identifying gene expression pattern changes based on co-expression analysis (Chaussabel and Baldwin 2014). The premise is usually to build networks of genes based on co-expression patterns, and to use this to group genes into modules, followed by testing for differential activity at the level of gene modules. The most famous example is weighted gene correlation network analysis, an implementation of such a method in R (Langfelder and Horvath 2008). The advantage of such co-expression methods compared to traditional differential testing is that differential activity can be tested at the level of gene modules, reducing the number of comparisons greatly and mitigating the loss of power due to multiple testing correction. In addition, the view of genes as parts of modules is a biologically appealing notion given that genes do not act alone; the expression of many transcripts should and do change in a coordinated fashion under different biological contexts (Chaussabel et al. 2008).



### **Pathway analysis**

While the strength of -omic analyses is such that changes across comparator conditions can be understood in an unbiased manner, the large volume of output can lead to difficulties in downstream biological interpretation of results. Gene lists produced by differential testing as above are not easily understood from manual inspection, but can instead be analysed systematically by looking for enrichment with known biological pathways. This is often done by comparing against annotated pathways e.g. as curated and maintained within the Reactome database (Fabregat et al. 2018) or by the Gene Ontology (GO) Consortium, or in the case of the Molecular Signatures Database (Liberzon et al. 2015) a combination of the former two and other curated pathways.

Statistically, enrichment can be detected either by over-representation testing with hypergeometric tests, or with leading edge analyses (Subramanian et al. 2005), for example implemented within the XGR (Fang et al. 2016) or ClusterProfiler (Yu et al. 2012) packages or other software.

### **Transcriptomic profiling at single cell resolution**

RNA-seq of samples as a whole represents the average mixture of RNA in the many different cells. The invention of single cell RNA-seq (scRNA-seq) in 2009 has further pushed the era of transcriptomics to the resolution of understanding the transcriptome of each individual cell (Tang et al. 2009). This has revealed significantly more heterogeneity in cells than previously understood, with new cell types or cell states constantly discovered as a result of the new technology.

One major analytical challenge presented by scRNA-seq is data sparsity, with only around 10% of values in the read count matrix per cell non-zero entries (Sarkar and Stephens 2021). This has often been referred to as a phenomenon of "dropout" or "zero inflation", with various models for scRNA-seq data utilising zero-inflated negative binomial (ZINB) distributions to account for this.

However, there is substantial debate on the degree to which zero-inflation is a technical artifact representing technical difficulties in RNA molecule capture (Kharchenko et al. 2014) versus biological realities such as transcriptional bursting (Svensson 2020; Jiang et al. 2022).

Aside from uncovering cellular heterogeneity, single cell profiling methods then allow comparison of differential cell type/state abundance, and DGE within the same cell type or state. These comparisons require new models and statistical frameworks beyond those discussed under bulk RNA-seq DGE. An in-depth consideration of them will be given in the relevant chapters.

### **Targeted transcriptomic profiling**

A major drawback of transcriptomic and especially single cell methods is experimental cost given the large volume of data. Alternatives include assaying a targeted panel of genes, either through qPCR methods or again through sequencing. A popular targeted scRNA-seq platform offered by Beckton Dickinson known as the Rhapsody is an example of the latter technology, with standard panels of 500 genes and up to over 1000 genes being profiled. This reduces sequencing costs and also mitigates issues with sparsity as discussed above, and has been used effectively in studying autoimmune conditions such as type 1 diabetes and SLE with CD4<sup>+</sup> T cell trajectories and heterogeneity uncovered (Trzupek et al. 2020).

## **1.6 Specific aims and objectives**

The overall goal of this thesis is to investigate sepsis pathophysiology through an integrated immunological and -omic approach. The following results chapters compare different sepsis etiologies including COVID-19 viral sepsis (chapters 3 and 4), influenza viral sepsis (chapter 4) and all-cause sepsis (chapters 3-5). This is done at differing cellular resolutions (bulk in chapter 3 and single-

cell in chapters 4-5, with only peripheral blood mononuclear cells in chapter 4 and whole blood in chapters 3 and 5) and with different modalities (RNA-seq and cell surface protein profiling either by sequencing (chapters 4 and 5) or mass cytometry (chapter 5)). Understanding pathophysiology behind the clinical syndrome will allow heterogeneity of the condition to be dissected, and specifically also within transcriptomic patient subgroups (sepsis response signatures, SRS) previously defined (Davenport et al. 2016; Burnham et al. 2017). These transcriptomic subgroups are discussed further in the relevant chapters.

The specific aims of this thesis are to:

- **Investigate the host bulk transcriptome in sepsis of different sources including viral sepsis due to COVID-19 (chapter 3).** The host peripheral blood bulk transcriptome in all-cause sepsis has mostly been interrogated either by microarrays or where using RNA-seq only the mononuclear cell fraction (Washburn et al. 2019). Further, the application of transcriptomic subphenotyping to COVID-19 and viral sepsis remains underexplored. In this chapter, I aim to:
  1. compare the host peripheral blood bulk transcriptome between all-cause non-COVID-19 sepsis and COVID-19 viral sepsis, and identify shared and distinct features across the two patient groups
  2. understand the applicability of previously defined transcriptomic subphenotypes (SRS) to non-ICU all-cause sepsis and COVID-19 viral sepsis cohorts
  3. elucidate the differences in gene expression response between septic and non-septic COVID-19 illness
- **Compare the peripheral blood mononuclear cellular transcriptomic landscapes of severe infection including in COVID-19, influenza, and all-cause sepsis (chapter 4).** The host response to severe infection involves a

coordinated immune response which we can examine through peripheral blood immune cell changes. Understanding the changes in cellular phenotype and transcriptomes will enable us to pinpoint specific cell types and states that are relevant in disease pathogenesis in severe infection. Specific aims are to:

1. build a multi-modal single-cell atlas to resolve cellular heterogeneity within the PBMC landscape, in the context of both the healthy state and under conditions of severe infection
  2. investigate the cellular compositional remodeling across severe infection in the context of COVID-19, influenza and all-cause sepsis
  3. characterize the gene expression changes per cell type across conditions of severe infection to understand the shared and distinct aspects of the immune response to different pathogens
- **Elucidate the cellular basis of transcriptomic subphenotypes in sepsis (chapter 5).** The heterogeneity in the sepsis response has necessitated strategies for stratifying patients into clinically and biologically meaningful subgroups. While subgroups e.g. SRS have been identified through peripheral blood transcriptomic subphenotyping, the molecular and cellular biological basis of these patient groups remains uncharacterized and the endotype status of these definitions controversial. Here, I aim to:
    1. generate a multi-modal single-cell atlas of all immune cells in the peripheral blood inclusive of granulocytes in the septic state
    2. elucidate the cellular compositional remodeling between sepsis SRS transcriptomic subphenotypes
    3. examine the cellular gene expression basis for SRS groups and thus identify identify cellular drivers of SRS
    4. validate the cellular drivers of SRS through integration of bulk RNA-seq

and mass cytometry data on separate sepsis samples

## MATERIALS AND METHODS

*This chapter describes the patient cohorts and recruitment, experimental procedures and analytical methods in this thesis.*

|                                                                 |    |
|-----------------------------------------------------------------|----|
| 2.1 Patient cohorts in chapters 3 and 4 . . . . .               | 34 |
| 2.2 Whole blood bulk RNA-seq . . . . .                          | 37 |
| 2.3 Peripheral blood mononuclear cell (PBMC) CITE-seq . . . . . | 39 |
| 2.4 Patient cohorts for chapter 5 . . . . .                     | 46 |
| 2.5 Whole blood scRNA-seq and Abseq . . . . .                   | 47 |
| 2.6 Neutrophil functional assays . . . . .                      | 53 |
| 2.7 Chapter 5 cohort 2/3 analyses . . . . .                     | 55 |

The methods for patient recruitment, data generation and data analysis for the results presented in chapters 3 and 4 have previously been published as part of the COMBAT consortium (COMBAT consortium 2022).

### 2.1 Patient cohorts in chapters 3 and 4

#### 2.1.1 Hospitalized COVID-19 patients

Patients were recruited into the Sepsis Immunomics (SI) study from the John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, UK, between March 13 and April 28 2020 if they displayed clinical features consistent with COVID-19 and tested positive for SARS-CoV-2 in an upper respiratory tract (nose/throat) swab specimen via reverse transcriptase polymerase chain reaction (RT-PCR). Ethical approval for SI was obtained from the South Central–Oxford

C Research Ethics Committee (REC) (REC reference 19/SC/0296). Whole blood was sampled from patients on days 1, 3 and 5 of hospital or intensive care admission.

### **2.1.2 Healthcare workers with COVID-19**

Healthcare workers based at Oxford University Hospitals NHS Foundation Trust, UK with COVID-19 symptoms and who tested positive on RT-PCR for SARS-CoV-2 in upper respiratory tract specimens (nose/throat swab) were recruited for a time of symptom matched set of community mild COVID-19 samples to the hospitalized COVID-19 patients. Recruitment was done under the Gastro-intestinal illness in Oxford: COVID substudy, with ethical approval from the Sheffield REC (REC reference 16/YH/0247].

### **2.1.3 Hospitalized patients with all-cause sepsis**

Non-COVID all-cause sepsis samples were also obtained under the SI study, from patients  $\geq 18$  years of age admitted to Oxford University Hospitals NHS Foundation Trust, UK with symptoms and signs of established sepsis (suspected infection with an acute change in total Sequential Organ Failure Assessment (SOFA) score of  $\geq 2$  points or a change in quick SOFA (qSOFA) score by  $\geq 2$  points). Patients were consented and recruited between 12 October 2019 and 13 March 2020. Blood sampling occurred on days 1, 3 and 5 of either hospital or intensive care admission.

### **2.1.4 Healthy volunteers**

Individuals 55 years or over and self-reporting as healthy with no symptoms of infection two weeks prior to approach were consented and recruited into the Genetic diversity and gene expression in white blood cells study [South Central Oxford REC B, reference 06/Q1605/55], with a single blood sample taken. The

donors were age and sex matched to the patient cohort.

### **2.1.5 Critically unwell patients with COVID-19 and influenza in London**

Influenza samples were taken from patients  $\geq 18$  years old on ventilatory support with an intensive care unit (ICU) stay  $\geq 24$  hours, with either a diagnosis of COVID-19 or severe influenza. Ethical framework was under the Aspergillosis in patients with severe influenza (AspiFlu ISRCTN51287266) study [Wales REC 5, reference 19/WA/0310] (Youngs et al. 2021).

Blood samples were taken from patients within 7 days of ICU admission, and within 72 hours of enrolment. Samples were obtained from three UK units including St George's University Hospitals NHS Foundation Trust, Guy's and St Thomas' NHS Foundation Trust and King's College Hospital NHS Foundation Trust. Diagnoses were confirmed by PCR.

### **2.1.6 The Sepsis Immunomics Study**

The inclusion and exclusion criteria for SI were as follows:

- Inclusion criteria:
  1. intensive care unit (ICU) patients with symptoms and signs of established sepsis (suspected infection with an acute change in total Sequential Organ Failure Assessment (SOFA) score of  $\geq 2$  points)
  2. emergency department / medical ward patients admitted with change in quick SOFA score by  $\geq 2$  points and a NEWS2 score of  $\geq 7$  or intensive care review requested
- Exclusion criteria (consistent with the GAINs study [NCT00121196]):
  1. patients or consultees unwilling or unable to give consent or advice



declaration

2. advanced directive to withhold or withdraw life sustaining treatment
3. admission for palliative care only
4. pregnancy and 6 weeks post-partum;
5. severe acquired immunodeficiency including systemic high dose steroid therapy (prednisolone 0.5mg/kg/day for 14 days or equivalent), HIV infection, known regular therapy with immunosuppressive agents e.g. azathioprine, or neutrophil counts <1000/mL due to any cause including metastatic disease and hematological malignancies or chemotherapy but excluding severe sepsis and solid organ/bone marrow transplant recipient receiving immunosuppressive therapy

## 2.2 Whole blood bulk RNA-seq

Whole blood from hospitalized Oxford patients (COVID-19 and sepsis), healthcare workers and healthy volunteers were sampled into Tempus tubes (Life Technologies Corporation) for extraction of whole blood total RNA for sequencing. RNA was extracted from Tempus tubes using the Norgen Preserved Blood RNA Purification Kit I (43400) and stored at -80°C before library preparation.

Total RNA-seq libraries were prepared by the Oxford Genomics Centre. After rRNA and globin depletion, libraries were produced with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. All 144 samples (124 patients) were sequenced in a single pool on one NovaSeq S4 flow cell (4 lanes), targeting 50 million 100bp read pairs per sample.

Due to limited material availability influenza samples were sequenced used a small-bulk RNA-sequencing method in order to facilitate genetic demultiplexing

of the CITE-seq data.

## **2.2.1 Bulk RNA-seq analysis**

### **Preprocessing**

Adaptor sequences were trimmed with TrimGalore (v0.6.2). Read mapping was performed with STAR v2.7.3 (Dobin et al. 2013) using the multi-sample 2-pass mapping approach against the reference genome GRCh38.100. Gene expression was then quantified with featureCounts (v1.6.4) (Liao et al. 2014) using Ensembl v100 annotations and genes filtered with a threshold of at least 10 reads in at least 10 samples, with 23,063 features retained for downstream analysis. RNA-seq data was normalized using the trimmed mean of M-values method R package from edgeR (Robinson et al. 2010).

### **Unsupervised exploratory analysis**

Log2-transformed, counts-per-million normalized data was used for principal component analysis (PCA) using prcomp (R v3.6.2) with default parameters. Hierarchical clustering analysis was performed with the hclust function (R v3.6.2) with Ward's method using the Euclidean distance metric as well as sample correlational distances. One sample was filtered out after PCA outlier status (Appendix), leaving n=143 samples (n=123 individuals) for downstream analysis.

### **Differential gene expression and pathway analysis**

Differential gene expression analysis on the normalized data was performed with linear mixed models as implemented in the limma R package (Ritchie et al. 2015). Genes were taken to be differentially expressed with thresholds of <0.05 for false discovery rate (FDR) and >1.5 for fold change. Pathway enrichment analysis was performed using Gene Ontology pathways via the clusterProfiler R package (Yu

et al. 2012) with Fisher's exact tests.

### **Sepsis response signature group assignment**

Sepsis response signature (SRS) assignments were performed by 1) assignment based on a logistic regression model trained on an in-house whole blood bulk RNA-seq dataset of 864 intensive care unit community acquired pneumonia or fecal peritonitis sepsis samples (from days 1, 3 or 5 of ICU admission), or 2) with the Sepstratifier package (Cano-Gamez et al. 2022). For the latter, the default options of k=20 nearest neighbours was used.

## **2.3 Peripheral blood mononuclear cell (PBMC) CITE-seq**

PBMCs were isolated from whole blood by density gradient centrifugation Leucosep tubes (Griener Bio-One) with Lymphoprep. PBMCs were stored in 10% DMSO + 90% fetal calf serum in liquid nitrogen until use.

To run all 140 samples for CITE-seq, samples were split into ten batches (pools) of 14 samples. Each batch contained at least one sample per comparator group as well as a similar age and sex distribution to minimize batch effects. PBMCs were thawed and washed, followed by pooling of all 14 samples per batch based on live cell counts to achieve the same number of live cells per sample. Viability sorting was then performed (7-AAD dye, BioLegend 420404) on a BD FACS Aria Fusion sorter.

After sorting, each pool was stained for cell surface proteins with a 192 TotalSeq-C antibody panel (BioLegend 99814) for 30 minutes on ice after FcX blocking (BioLegend) for 10 minutes on ice first. Cells were counted on the BioRad TC20 Automated Cell Counter and loaded onto the 10X Genomics Chip G at 50,000 cells per channel. Each pool was loaded across seven channels. From each

channel, four libraries were generated, gene expression (GEX), surface proteome (ADT) as well as T and B cell receptor repertoires. The libraries were sequenced on an Illumina NovaSeq 6000 sequencer across nine S4 flowcells (4 lanes per flow cell).

### 2.3.1 Analysis

#### Pre-processing

FASTQ files were generated with CellRanger version (v3.1.0) with the mkfastq command using tge GRCh38 reference and Ensembl version 100 annotations and gene expression and ADT matrices produced with the CellRanger count command. Cell calling was performed with the union of cells identified by a combination of CellRanger version (v3.1.0) and EmptyDroplets (v1.8.0) (Lun et al. 2019). Samples in each pool were genetically demultiplexed with Vireo (v0.4.0) (Huang et al. 2019) using the genotypes from the bulk RNA-seq data.

Sequencing quality was high with a sequencing saturation of 77-82%, >85% of RNA reads >Q30 and >85% reads mapped to the genome (Appendix). Cells were then filtered for high mitochondrial reads ( $\geq 10\%$ ) and a low number of genes ( $< 300$ ) to remove dead or dying cells and doublets respectively.

To prepare for downstream cell clustering and cell type and state identification, the full dataset was log normalized with 3000 highly variable genes (HVG) selected in scanpy (flavor = 'seurat\_v3') (Wolf et al. 2018). The HVG selection process excluded a list of HLA-genes as well as genes coding for the variable regions of lymphocyte receptors in order to avoid spurious clustering from differential mapping of HLA genes to the reference for different samples or clustering of lymphocytes purely by receptor gene usage.

The HVGs were used downstream for PCA. Batch correction was attempted with python implementations of bbknn, scanorama and Harmony, with UMAP

visualizations showing superior mixing of pools with Harmony. Final dataset alignment was therefore performed with the PCs batch corrected with Harmony using the top 65 PCs with the sample ID (i.e. 140 distinct levels) as the batch variable.

For the ADTs, each channel was normalized independently with the denoised and scaled by background (DSB) algorithm (Mulè et al. 2022). The cells identified from the GEX data as the foreground and unassigned droplets with  $\log_{10}(\text{GEX\_nUMI} + 1) \geq 1.5$  as the background.

### **Clustering and multimodal annotation**

Multimodal annotation was performed with expert immunological knowledge from the COMBAT consortium to guide a curated integration of the data from the different modalities for cell type/state annotation.

GEX and ADT data were clustered separately. The ADT surface phenotype was prioritized for defining major cell lineages and subsets first before GEX based clustering. Cell types and subsets were further refined using information from the repertoire and GEX layers.

The normalized ADT data was subject to hierarchical stochastic neighbor embedding (hSNE) and unsupervised Gaussian mean shift (GMS) clustering as implemented in Cytosplore (Höhl et al. 2016). This analysis, which was based on the expression of all n=192 ADT tags, partitioned the dataset into 66 discrete clusters. Based on expression of known lineage markers (CD3, CD4, CD8, CD56, CD19, CD20, CD27, CD38, CD14, CD123, CD1c, CD33, CD235, CD34, V $\delta$ 2, V $\alpha$ 72 and CD161) these clusters were assigned as NK cells, B cells, plasmablasts, V $\delta$ 2+ T cells, CD14+ Monocytes, CD123+ pDC, CD1c+ DC, red blood cells (RBC), CD34+ cells and distinct populations of CD3+ T cells. After exclusion of likely contaminating CD33+ CD3+ and CD56+ cells within the B cell and plasmablast compartment, the B cell and plasmablasts were reclustered based

on the expression of CD10, CD24, CD25, CD19, CD20, CD22, HLADR, CD21, CD23, IgD, CD1c, IgM, CD38, CD29, CD71, CD39, CD27, HLA-ABC and IgA.

In total, this analysis of the ADT data identified 12 T cell subpopulations, 7 B cell/plasmablast subpopulations and five distinct sets of doublets (with markers of more than one major immune lineage). These clusters were intersected with information from the GEX and repertoire modalities for the final multimodal annotation of the dataset as described below.

After identifying ADT lineage clustering analysis, cells were subsetted by lineage, with variable gene selection and alignment redone with Harmony. Nearest neighbor graph construction was done with Euclidean distance and the 'hns' algorithm (Bergen et al. 2020) followed by Leiden clustering in scanpy. Gene marker discovery for each cluster was done with Wilcoxon rank-sum tests in Seurat.

In the first step the major cell types – T/NK, myeloid and B/plasmablast cells were extracted into n=3 separate subsets. Each of these three subsets was then separately aligned and clustered (45 PCs for n=482k T/NK cells; 45 PCs for n=279k myeloid cells; 40 PCs for n=66k B/plasmablast cells). Iterative variable gene selection, alignment and clustering was repeated separately on six subsets: (A) CD4+ T cells (50 PCs, n=302k cells), (B) CD8+ T cells (45 PCs, n=180k cells), (C) myeloid cells (50 PCs, n=261k cells), (D) B cell/Plasmablast cells (PCs, n=57k cells), (E) doublets (40PCs, n=26k cells), and finally (F) other cells (50 PCs, n=10kcells).

The six cells subsets were clustered at a range of resolutions (1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 6, 8). From a base resolution for each subset, sub-populations were refined based on expert knowledge of cell types and states by splitting or merging clusters at higher or lower cluster resolutions, leaving a final total of n=131 GEX clusters. All three modalities were combined to identify doublets, with congruence required between GEX Scrublet scores, ADT cell surface phenotype

and immune receptor sequences.

After intersecting GEX, ADT and repertoire information, n=343 clusters (n=782,245 cells) of singlet cell clusters were retained for downstream analysis. Each of these clusters were then assigned to a “minor subset,” “major subset” and “cell type” category for analysis at different levels of granularity.

### **Cell composition differential abundance (DA) analysis**

Six samples were excluded from composition analysis due to low cell numbers (three with < 500 cells per sample and three samples with confirmed or suspected malignancy). For hospitalized patients (COVID-19 and sepsis), only samples closest to maximum severity were retained i.e. only one sample per individual used. The final number of samples analyzed was therefore: healthy volunteers, n=10; COVID-19 acute in-patient mild, n=12; COVID-19 acute in-patient severe, n=20; COVID-19 acute in-patient critical, n=18; COVID-19 community COVID-19, n=12; influenza, acute in-patient, n=10; and sepsis acute in-patient, n=15.

Composition analysis was performed for the different levels of cellular granularity, including for minor cell subsets as well as individual high resolution immune clusters in the T/NK cells, B/plasmablast clusters and mononuclear phagocyte (MNP clusters). Percentages were calculated out of total PBMCs for minor subsets, while the denominators for higher resolution immune cell clusters were total T and NK cells, total B and PB cells, or total MNPs.

Quantification of statistical significance of differential cell subset/cluster abundance across comparator groups was done with edgeR (v3.28.1) by modeling cluster counts using a quasi-likelihood negative binomial generalized log-linear model (glmQLFit function). Age, sex and pool were included as covariates. Clusters with fewer than ten cells in under 10 individuals were filtered out. Counts were normalized by the total number of cells in each sample for the major and minor cell subset analyses, or by the total myeloid, the total B and

plasmablast, or the total T and NK cells for higher resolution immune cell cluster analyses.

To determine the statistical significance of differences in cell subset/cluster abundance between groups, differential abundance analysis was performed using edgeR, v3.28.1 (Amezquita et al., 2020). Cell subset/cluster counts were modeled adjusting for age, sex and sample pool using a quasi-likelihood negative binomial generalized log-linear model (glmQLFit function), with pre-filtering to remove clusters with a count of  $n \leq 10$  cells in  $\leq 10$  individuals. FDR was used to adjust for multiple comparisons based on the number of clusters and pairwise condition comparisons. For the hospitalized COVID-19 patients, edgeR analysis was performed to identify statistically significant associations between cluster abundance and experimental or clinical variables with ANOVA tests.

### **Weighted gene correlation network analysis (WGCNA) and pathway analysis**

For each of  $n=7$  "major cell types" i.e. B cells, Plasmablasts, NK cells cMono, ncMono, CD4 T cells and CD8 T cells, WGCNA was performed for gene expression analysis. Sample pseudobulked counts were used, with exclusion of the same samples as in differential abundance analysis.

The genes used for WGCNA were filtered for each cell type based on a certain threshold in a given minimum number of samples. Based on inspection of expression distributions, thresholds of counts per million of 1 and 3 were chosen for cMono, NK cells, CD4 T cells and CD8 T cells vs. ncMono, PB and B cells respectively. The minimum number of samples was based on the smallest patient group i.e. 10 (healthy volunteers and influenza). After filtering, 9,000-12,000 genes per "major cell type" was used for WGCNA. Count matrices were then normalized per million (cpm) and  $\text{cpm}+1$  log transformed before being batch corrected by pool with the ComBat algorithm (sva R package v3.36.0) (Leek et al. 2012).



The normalized and batch corrected count matrices were used to build a signed-hybrid gene co-expression network with biweight midcorrelation. Genes were clustered into modules, and modules with eigengene (the first principal component of a module) dissimilarity  $\geq 0.3$  were merged. This left a total of 64 modules. Three “ambient\_rna” modules which had strong eigengene correlation with ambient RNA abundance were excluded (Spearman’s  $\rho \geq 0.83$ ,  $P \leq 2.2 \times 10^{-16}$ ). Unassigned grey WGCNA modules which were discovered in each cell type except NK cells (total = 6 grey modules) were also excluded from downstream analysis, leaving a total of 55 modules.

Pathway analysis was done by over-representation analysis using the gsfisher R library <https://github.com/sansomlab/gsfisher>. Modules were then annotated based on by inspection of gene membership, enrichment of biological pathways (GO Biological Process, GO Cellular Component, KEGG, MSigDB REACTOME and MSigDB transcription factor motifs) and custom gene sets. The “curated type I IFN response” geneset included *GBP2*, *IFI27*, *IFI44L*, *IFIT1*, *IFITM1*, *IFITM3*, *IFNA1*, *IFNA10*, *IFNA13*, *IFNA14*, *IFNA16*, *IFNA17*, *IFNA2*, *IFNA21*, *IFNA4*, *IFNA5*, *IFNA6*, *IFNA7*, *IFNA8*, *IFNB1*, *IRF1*, *IRF3*, *IRF7*, *IRF9*, *ISG15*, *MX1*, *MX2*, *OASL*, *RSAD2*, *SIGLEC1*, *TRIM56*, *USP18* and *XAF1*. The “curated AP1 TF family” geneset comprised of *ATF3*, *ATF4*, *ATF5*, *ATF6*, *FOS*, *FOSB*, *FOSL1*, *FOSL2*, *JUN*, *JUNB*, *JUND*, *MAFB*, *MAFF* and *MAFK*. The custom “zf-C2H2 (PF00096) domain” geneset consisted of PF00096 domain containing genes (Pfam database). Separate FDR corrections were performed within major cell type and ontology.

Module eigengene disease associations were computed as correlations, coding the given disease group contrasts 1 and 0, while associations with variance between patient groups were tested by ANOVA (COVID+HV, All Groups ANOVA). Pairwise-complete observations were used for correlations between module eigengenes and quantile normalized clinical variables.

## 2.4 Patient cohorts for chapter 5

### 2.4.1 Cohort 1

Healthy individuals (by self-reporting) (healthy volunteers, HVs) with no clinical features of infection in the past 2 weeks were consented and recruited into the Genetic diversity and gene expression in white blood cells study [South Central Oxford REC B, reference 06/Q1605/55]. Blood samples were collected on one occasion.

Sepsis and post elective cardiac surgery patients were consented into the Sepsis Immunomics Study [South Central Oxford REC C, reference:19/SC/0296] between May and November 2021. Samples were collected on days 1, 3 or 5 of either hospital or intensive care admission. Written informed consent was obtained from adults or advice declarations from personal/nominated consultees for patients lacking capacity, with retrospective consent obtained from the patient once capacity was regained.

Patient sepsis samples were collected in the acute phase of illness from adult patients ( $\geq 18$  years of age) admitted to Oxford University Hospitals NHS Foundation Trust, UK. Patients were recruited either from the ICU if they had symptoms and signs of established sepsis (suspected infection with an acute change in total Sequential Organ Failure Assessment (SOFA) score of  $\geq 2$  points), or from the emergency department and medical wards if they had a change in quick SOFA score by  $\geq 2$  points and a NEWS2 score of  $\geq 7$  or intensive care review requested. Convalescent sepsis patient samples were collected from individuals with acute sepsis samples already taken 1-6 months after hospital discharge.

Post elective cardiac surgery samples were collected from adult patients ( $\geq 18$  years of age) admitted to Oxford University Hospitals NHS Foundation Trust, UK. Patients were deemed eligible if they were 1) undergoing cardiac bypass

surgery, 2) required post-operative intensive care unit stay and 3) did not have an infection prior to surgery. Exclusion criteria were identical to those for sepsis patients.

### **2.4.2 Cohort 2**

HVs and sepsis patients were as described above for the patient cohorts as part of the COMBAT consortium.

### **2.4.3 Cohort 3**

Adult patients ( $\geq 18$  years old) were recruited through the UK Genomic Advances in Sepsis (GAINs) study (NCT00121196) upon admission to intensive care with sepsis due to community-acquired pneumonia (CAP) or faecal peritonitis (FP); (<https://ukccggains.com/>). Patients were recruited from 34 UK ICUs, with diagnoses, inclusion and exclusion criteria as described by Davenport et al (2016). Ethics approval was granted nationally and locally (RECs 05/MRE00/38, 08/H0505/78, and 06/Q1605/55), with informed consent obtained from all patients or their legal representative.

## **2.5 Whole blood scRNA-seq and Abseq**

### **2.5.1 Whole blood single cell RNA and cell surface protein profiling (cohort 1)**

Blood samples were taken into EDTA Vacutainer tubes (BD Biosciences) and processed within sixty minutes of collection. Red blood cell lysis was performed by mixing one mL of whole blood with nine mL of 1X eBioscience lysis buffer (Thermofisher) for 10 minutes. The RBC lysis was repeated once for three minutes with with three mL lysis buffer and identical quenching. 200,000 cells were transferred for antibody staining and single cell capture.

Single cell RNA and cell surface protein sequencing was performed with the BD Rhapsody platform (633731 and 633733) according to manufacturer's recommendations. The whole transcriptome amplification assay (633801) and 30 Rhapsody Abseq antibodies (Appendix) were used. A capture of 6000 cells per sample was aimed for according to manufacturer's dilution instructions. Manufacturers' recommendations were followed for reverse transcription, complementary DNA (cDNA) amplification (633773) and library construction (633801) in six batches. Libraries were sequenced on the NovaSeq6000 (Illumina).

### **2.5.2 Whole blood single cell multi-omic analysis (cohort 1)**

Analysis was performed with v4.0.0 R and python 3.8.6.

#### **Preprocessing and quality control**

Read mapping was performed with STARsolo (v2.7.9a) using the GRCh38 reference genome. Unfiltered files with all 884,736 cell barcodes on the whitelist were called downstream for cells by the emptyDrops function in DropletUtils (v1.10.3) with a UMI threshold of 100 and FDR of 0.5%. Spliced and unspliced counts were quantified with the Velocyto option in `-soloFeatures`.

Abseq reads were trimmed with Trimmomatic (v0.39) to 48 bases i.e. 12bp UMI + 36bp Abseq nucleotide sequence and mapped to an artificial reference of the Abseq nucleotide sequences also using STARsolo.

Cells expressing <100 genes, >4000 genes, >10% mitochondrial reads, >2% hemoglobin reads and with a  $\log_{10}(\text{UMI}/\text{gene}) < 0.6$  were removed. Genes expressed in <10 cells or with a total count of <3 were removed (Appendix).

Scrublet (v.0.2.3) and doublet detect (v3.0) were both used on default settings to generate doublet scores.

**Normalization, dimensionality reduction, visualisation and clustering**

Data were normalized (scanpy v1.7.2: `normalize_total`), log+1 corrected (scanpy: `log1p`) and 4000 HVGs identified using the Seurat `vst` algorithm (scanpy: `highly_variable_genes`, `flavor='seurat_v3'`).

The TotalVI variational autoencoder framework (scvi-tools v0.10.0) (Gayoso et al. 2021) was used on default settings for multi-modal dimensionality reduction. All 30 proteins and HVGs were used, and batch correction was performed with each individual sample as all samples were collected and processed individually. The resulting 20 latent dimensions were input for downstream UMAP visualization and graph construction for clustering and differential cell abundance analysis. Unsupervised clustering was done in Seurat (v4.0) with  $k=30$  nearest neighbours. Choice of  $k$  hyperparameter for number of nearest neighbours was made following inspection of Clustree (v0.4.20) analysis of a range of values for  $k$  and the effect on cluster splitting. The smart local moving algorithm was used for clustering in the Seurat `FindClusters` function (`algorithm=3`).

Clustering resolution assessment was done by combining three metrics including evaluation of cluster average silhouette width, cluster neighbourhood purity, and a 30 iteration bootstrap to determine cluster stability with respect to sampling noise (bluster v1.0). One additional expert driven element was employed i.e. inspection the top defining markers of clusters at varying resolutions to match clustering results with known biology and understand potential value in merging vs. splitting clusters. Combining these four measures in total, the default clustering resolution of 0.8 was chosen, with the splitting of one immature neutrophil cluster into two (*MPO*<sup>+</sup> immature neutrophils/progenitors and *PADI4*<sup>+</sup> immature neutrophils), based on the clustering resolution of 1.1.

**Cell type and state annotation and RNA velocity/trajectory analysis**

Cell types and states as defined at the above chosen resolutions were merged for annotation of major known immune cell types at a broad level, and kept at the clustering resolution of choice for annotation at a fine granularity. Cell surface protein markers were used to define major cell lineages which were merged for the broad level annotations. This assignment was confirmed by SingleR (v1.6.1) automatic assignments.

Fine annotation was then performed based on inspection of biologically known/potentially meaningful gene markers to subdivide T cell and neutrophil populations. For neutrophil clusters, maturity annotation was assisted by CD10 expression (a marker of neutrophil maturity) on UMAP projection. Clusters were annotated according to gene markers that particularly distinguished the neutrophil state.

The unspliced and spliced counts were used for RNA velocity analysis of the neutrophils (without degranulating or apoptosing neutrophils) with scVelo (v0.2.3, default settings of stochastic model) (Bergen et al. 2020). The TotalVI latent dimensions were used for moments estimation, and the annotated neutrophil clusters were used for partition based graph abstraction.

**Differential abundance (DA) analysis**

Cell type/state DA changes across comparator groups was tested by sampling neighborhoods of cells from a k-nearest neighbors graph and looking for enrichment of either condition in each neighborhood as implemented in MiloR (Dann et al. 2021). All 20 batch corrected TotalVI latent dimensions were used for MiloR (v0.99.19) kNN graph construction (k=30). 10% of cells were sampled for neighbourhood indexing. Differential testing was then performed with generalized linear models including age and sex as covariates. Significantly different neighbourhoods were determined by a spatial corrected FDR <0.05.

**Differential gene expression (DGE) analysis**

Gene expression per individual and for each cell type/states were aggregated into a pseudobulk count matrix i.e. raw UMI counts per gene were summed across all cells belonging to the same cluster and to the same individual. Pseudobulk gene filtering was done based on minimum expression of  $n$  counts in at least  $X$  samples, where  $X$  was the smallest comparator group, while  $n$  was chosen per pseudobulk based on histograms of logged count distributions to filter out lowly expressed genes which would contribute to noise. Pseudobulks were edgeR TMM normalized (v3.30.3) and log transformed for PCA. Pseudobulk DGE was done with edgeR generalized linear models including age, sex and sequencing batch as covariates.

**Consensus non-negative matrix factorization (cNMF) of neutrophil gene expression**

First, 1000 highly variable genes per neutrophil state were obtained with the Seurat function `SelectIntegrationFeatures`. Each of the neutrophil states (except degranulating neutrophils, apoptosing neutrophils and cycling neutrophil progenitors) were then analyzed with cNMF (v1.2) as previously described (100 factorizations) (Kotliar et al. 2019). Factorizations were done for four to 15 programs, with final program number for each neutrophil state chosen based on a tradeoff of stability vs. error for a range of program numbers factorized out as recommended by the original authors. Mean program usage per sample for each neutrophil state was correlated with SRSq scores per sample with FDR adjustment for all programs except programs which picked up outlier individual cell-specific expression signatures (*IL1R2*+ immature neutrophil program 1, *PADI4*+ immature neutrophil program 5, *MPO*+ immature neutrophil/progenitor program 7). The top 50 genes per program were taken for downstream pathway and transcription factor prediction analysis.

### **Transcription factor (TF) prediction analysis**

Gene programs positively correlated with SRSq were used for TF prediction analysis with the Cytoscape (v3.9.1) plugin iRegulon (v1.3) (Janky et al. 2014). Prediction was done with the top 50 genes of each of the gene programs. All analysis was performed on default settings. TFs with a normalized enrichment score (NES) >3.0 were extracted. An identical TF analysis was performed for module 10 genes from cohort 2 RNA-seq WGCNA analysis (below) with the top 1% genes correlating with the module eigengene.

### ***IL1R2*+ neutrophil defining gene set**

A gene set specific to *IL1R2*+ neutrophils was identified by taking the gene markers of *IL1R2*+ neutrophils from sepsis samples and filtering for an adjusted P-value of <0.01 and an average log<sub>2</sub>FC >1 to obtain markers specific to the cells.

### **Gene set enrichment scoring for single cells and pathway enrichment analysis**

Neutrophils were scored for expression of gene sets defining granulocytic myeloid derived suppressor cells (G-MDSC) (Bayik et al. 2020; Alshetaiwi et al. 2020) with the AddModuleScore function in Seurat, followed by two sided Wilcoxon rank-sum tests to test for differences across different groups of cells.

Pathway enrichment analysis was performed against GO and MSigDB Hallmark pathways via the ClusterProfiler R package (v4.1.4) with over-representation analysis via Fisher's exact tests as implemented in the enricher function.

### **Sepsis endotype (sepsis response signature) assignment/scoring**

SRS endotypes were assigned for sepsis samples by aggregating gene expression per individual for all cells to produce a per sample pseudobulk count matrix. Genes were filtered per sample pseudobulk based on minimum expression of five counts in at least X samples, where X was the smallest sample group by



comparator group i.e. six HVs. Sample pseudobulks were then normalized by the EdgeR trimmed mean of M-values (TMM) method. The relevant genes (*DYRK2*, *CCNB1IP1*, *TDRD9*, *ZAP70*, *ARL14EP*, *MDC1*, *ADGRE3*) were extracted for SRS assignment and SRSq score calculation by Sepstratifier (v0.0.0.9) (Cano-Gamez et al. 2022).

## 2.6 Neutrophil functional assays

### 2.6.1 Isolation of neutrophils from whole blood

5-10 mL whole blood from EDTA Vacutainer tubes (BD Biosciences) was used with the EasySep HLA Chimerism Whole Blood CD66b positive selection kit (StemCell) following manufacturer's instructions. Isolated neutrophils were washed with HBSS (Lonza) and resuspended in complete media (RPMI-1640 + 10 fetal calf serum (FCS), 100 IU/mL L-glutamine, 100 IU/mL penicillin-streptomycin) for downstream assays.

### 2.6.2 Phagocytosis Assay

A 20 minute incubation of neutrophils with pHrodo Green E.Coli Bioparticles (Invitrogen) was done for assessing phagocytosis. Beads were mixed with cells at a concentration of 1 in 10. A negative control with no beads was done. Cells were then washed and stained for surface markers for 30 minutes with 7-AAD, CD66b-AF700 (G10F5) and Siglec-8-APC (7C9) antibodies from Biolegend and fixed using Fix/Lyse solution (Invitrogen) for 10 minutes before flow cytometric acquisition using a BD LSRFortessa X-20 analyzer. The fluorescence minus one (FMO) control sample that was not incubated with any pHrodo bioparticles was used for median fluorescent intensity (MFI) calculation of gated 7AAD- CD66b+ Siglec-8- singlets i.e. MFI of control sample was subtracted from MFI of test sample for the final MFI of the test sample.

### 2.6.3 Neutrophil - allogeneic CD4<sup>+</sup> T cell co-culture

Healthy donor leukocyte cones were used for CD4<sup>+</sup> T cell isolation and cryopreservation. The T cells were thawed and stained for use with eBioscience Cell Proliferation Dye eFluor450 (Invitrogen) at a final concentration of 10  $\mu$ M for 20 minutes. T cells were co-cultured with neutrophils isolated from whole blood in 96 well U bottom plates at 37°C 5% CO<sub>2</sub> at a 4:1 neutrophil to T cell ratio. Cultures were done in complete media supplemented with 50 IU/mL recombinant human IL-2 (Biolegend) and with anti-CD3/CD28 dynabeads (Gibco). Anti-CD3/CD28 beads were added for a 1:1 bead to T cell ratio and the total cell density was maintained at 200,000 cells per well. As controls, T cells were plated alone without neutrophils in every run, with and without anti-CD3/CD28 beads.

To reverse T cell inhibition, 1 mM L-arginine (Sigma-Aldrich), 1  $\mu$ M arginase-1 inhibitor CB-1158 (Fisher Scientific), 10 ng/mL anti-CD274 (eBioscience clone MIH1), 10 ng/mL anti-CD273 (PD-L2) (eBioscience clone MIH18), 20  $\mu$ g/mL PGE2 (Sigma-Aldrich), 10  $\mu$ M indomethacin (Sigma-Aldrich), 15  $\mu$ M EP2 inhibitor TG-6-10-1 (MedChemExpress) or 15  $\mu$ M EP4 inhibitor GW627368 (MedChemExpress) were added to co-cultures.

After 72-96 hours of co-culture, cells were stained for surface markers for 30 minutes at room temperature, followed by Annexin V (apoptosis) and 7-AAD (live/dead) stains for 15 minutes. Samples were analyzed using the BD LSRFortessa X-20. All antibodies were purchased from Biolegend unless otherwise stated: CD3-APC (UCHT1), CD4-BUV395 (BD Biosciences, SK3), CD66b-AF700 (G10F5), PD1-PE (NAT105), CD69-PECy7 (FN50), 7-AAD, Annexin V- FITC. T cells were identified as a live single cell CD66b<sup>-</sup> CD4<sup>+</sup> CD3<sup>+</sup> Annexin V<sup>-</sup> 7AAD<sup>-</sup> population. Proliferation analysis by dye dilution was established using the non-bead stimulated T cells cultured alone without

neutrophils as the baseline for determining the proliferative fraction of cells in each sample without stimulation. The proliferative fraction and percentage of cells expressing PD-1 and CD69 were calculated as a percentage relative to the anti-CD3/28 bead stimulated no co-culture T cells control to account for donor variation.

### **2.6.4 Flow cytometry data analysis**

Flow cytometry data was analyzed using Flowjo (v10) and Graphpad Prism (v9).

### **2.6.5 NETosis assay**

Neutrophils were plated at 20,000 cells/well in low auto-fluorescence F-12K Nut Mix (Gibco) media (Hams-F12) using flat-bottomed 96-well plates previously coated for 1 hour with 0.01% poly-L-ornithine solution (Sigma). Cells were allowed to attach for up to 1hr prior to treatment with 100nM phorbol 12-myristate 13 acetate (PMA), as per manufacturer's recommendations, or left untreated as controls. 250 nM IncuCyte Cytotox Green Dye was added to measure NETosis-induced cell death. The final volume in each well was made up to 100  $\mu$ L with Hams-F12 media. The plated cells were imaged with the IncuCyte Live-Cell Analysis System incubator every 20 minutes for up to 4 hours, using standard scan type, objective 20X with green fluorescence channel selection.

## **2.7 Chapter 5 cohort 2/3 analyses**

### **2.7.1 Bulk RNA-seq analysis (Cohort 2)**

Analysis was performed with v4.0.0 R. The edgeR TMM normalized and filtered count matrix was obtained from the COMBAT consortium (COMBAT consortium 2022), with 23,063 genes from 42 samples analyzed.

For cohort 2 sepsis samples, the relevant genes (*DYRK2*, *CCNB1IP1*, *TDRD9*,

*ZAP70, ARL14EP, MDC1, ADGRE3*) were extracted from the normalized count matrix and used for SRS assignment and SRSq score calculation by Sepstratifier (v0.0.0.9). PCA was performed with the top 15% most variable features after log transformation.

WGCNA (v1.70-3) was conducted with the log normalized count matrix. The parameters used were a soft thresholding power of 4 which was the lowest power reaching a scale free topology model fit of 0.8, a signed-hybrid network, biweight midcorrelation and a deepsplit parameter of 2 with minimum module size of 30. Modules were summarized by eigengene expression (the value of the first principal component of the module) and modules with eigengene correlation of  $\geq 0.2$  were merged for a total of 33 modules identified. Differential testing of modular eigengene expression across SRS groups was conducted with two-sided Wilcoxon rank-sum tests, adjusting for multiple testing with the FDR.

Pathway enrichment analysis was done with the ClusterProfiler R package. Over representation analysis with Fisher's exact tests as implemented in enricher function was used to test for enrichment of neutrophil gene sets from the Human Cell Atlas (Hay et al. 2018) and signatures of differentiating and differentiated neutrophils from mouse single cell transcriptomics (Giladi et al. 2018). Gene set enrichment analysis for cohort 1 *IL1R2*<sup>+</sup> neutrophil defining genes was performed via the GSEA function for the WGCNA module most correlated with immature neutrophil proportions.

### **2.7.2 CyTOF analysis (cohort 2)**

Analysis was performed with v4.0.0 R.

#### **Pre-processing**

CyTOF FCS data files were obtained from the COMBAT consortium (COMBAT consortium 2022) after normalization, debarcoding, and bead and doublet

removal. To avoid biases due to highly varying cell numbers per sample, a maximum of 50,000 cells per sample was taken. As not all batches contained control samples for batch correction, a control sample free correction method Harmony (v1.0) (Korsunsky et al. 2019) was adapted from scRNA-seq methodology for use given previous demonstration of utility in CyTOF data (Ogishi et al. 2021).

### **Clustering, trajectory inference and differential abundance analyses**

The batch corrected cell by marker intensity matrix was clustered with FlowSOM (Quintelier et al. 2021) as implemented in the CATALYST package (v1.14.0) (Nowicka et al. 2019). This was first done with all cells and all 41 markers (12x12 starting grid). Consensus clustering was then used for cluster merging to 50 metaclusters which were then manually annotated and further merged. As the 41 marker panel was designed to be myeloid focused i.e. to capture both monocyte and neutrophil subsets, five separate monocyte clusters were annotated. Fine neutrophil annotation was then done by subsetting the neutrophils and reclustering to 30 metaclusters with 17 selected markers. These neutrophil clusters were again manually annotated and merged based on median marker expression, with marker enrichment modeling scores (Diggins et al. 2017) confirming identity of the eight annotated subsets. This resulted in a final total of 22 clusters at the finest level of annotation (14 non-neutrophil and eight neutrophil).

For trajectory analysis, neutrophils were downsampled to 15,000 cells per comparator group (SRS1, non-SRS1 and HVs) for diffusion map dimensionality reduction as implemented in the CATALYST package followed by principal curve fitting on the minimum spanning tree with default settings on the slingshot function in the Slingshot package (v1.8.0) (Street et al. 2018).

For DA analysis, the number of cells identified in each of the 22 clusters was

divided by the total number of cells in the patient to evaluate the proportion of the cluster. PCA of cluster proportions was done on quantile normalized cell proportions. Differential testing of cluster abundances was performed using a generalized linear mixed model using the `glmmPQL` function from the `MASS` package (v7.3-51.6) to account for repeated sampling in six patients. Batch was included as a covariate in the model and patient included as a random effect, with an extra random effect term for each individual sample also included to model the overdispersion in proportions seen in high dimensional cytometry data (Nowicka et al. 2019). Adjustment for multiple testing of each cluster was performed with the FDR.

### **2.7.3 Multi-modal analysis by MOFA+ (cohort 2)**

MOFA+ (multi-omic factor analysis) (v1.0.1) (Argelaguet et al. 2020) was used to integrate cohort 2 bulk RNA-seq and CyTOF data. MOFA+ identifies latent spaces of shared variation across datasets. As a negative control, a plasma proteomic dataset from the same patient samples was included (chapter 5). Concentrations of 105 plasma proteins as determined by the `timsTOF` mass spectrometer were obtained from the COMBAT consortium for this (COMBAT consortium 2022). For each sample, the eigengenes of the 33 identified modules from WGCNA, quantile normalized proportion of the 22 cell clusters from CyTOF clustering and annotation, and  $\log_2$  transformed and median centered intensity values for 105 plasma proteins were utilized, with six samples missing plasma protein data and one sample missing CyTOF data. Views were left unscaled, number of factors was set at 10, and all other model options were left with default values. Differences in factor values between SRS groups was tested with a two-sided Wilcoxon rank-sum test, with FDR adjustment.

#### **2.7.4 Cytokine analysis (cohort 2)**

Concentrations of 51 plasma analytes as determined by the Luminex assay were obtained from the COMBAT consortium. Differential testing of cytokine concentrations across SRS groups was conducted with two-sided Wilcoxon rank-sum tests, with FDR adjustment.

#### **2.7.5 Cell type/state deconvolution of 864 sepsis bulk RNA-seq samples (cohort 3)**

Bulk RNA sequencing data were obtained for 864 samples (667 patients) from the GAIN study (EGAD00001008730). Briefly, whole blood samples were taken on the first, third, and/or fifth day following ICU admission and leukocytes were isolated using the LeukoLOCK system (Thermo Scientific). RNA was extracted using the manufacturer's protocol (Ambion), and libraries prepared using NEB Ultra II Library Prep kits (Illumina) were sequenced on a NovaSeq6000 (Illumina). Reads were aligned to the GRCh38 reference genome using STAR (v2.7.3) and gene expression quantified using featureCounts. Finally, the count matrix was normalized using TMM from edgeR and log-transformed.

Cell type/state abundance deconvolution was then done with CIBERSORTx (Newman et al. 2019). First, a reference signature matrix was constructed with the cohort 1 single cell dataset, with each finely annotated cell type and state except apoptosing cells downsampled to 100 cells per type/state. The signature matrix was created with the Create Signature Matrix analysis module with parameters min.expression=0.25, replicates=100 and sampling=0.5. Cell fractions of the 864 sample whole blood bulk RNA-seq dataset were then estimated with the Impute Cell Fractions analysis module using the single cell reference matrix, with batch correction S-mode enabled and quantile normalization disabled. For DA testing of cell types/states across SRS, only the

667 samples from unique individuals were used for two-sided Wilcoxon rank-sum tests per cell type/state with FDR adjustment.



## THE BULK TRANSCRIPTOMIC RESPONSE TO ALL-CAUSE AND COVID-19 SEPSIS

*This chapter explores the host peripheral blood bulk transcriptomic response in COVID-19 and all-cause sepsis.*

|     |                        |     |
|-----|------------------------|-----|
| 3.1 | Introduction . . . . . | 61  |
| 3.2 | Results . . . . .      | 66  |
| 3.3 | Discussion . . . . .   | 98  |
| 3.4 | Conclusions . . . . .  | 107 |

### 3.1 Introduction

Sepsis is a heterogeneous clinical syndrome with myriad microbial etiologies, with no approved molecularly targeted therapies based on a rational understanding of the immune state of the individual. One major reason for this lack of treatments is the insufficient understanding of the pathogenesis of the condition and the heterogeneity of it.

In this chapter, I describe the use of bulk peripheral blood transcriptomic data to interrogate these gaps in our understanding, both of all-cause sepsis and in a specific form of viral sepsis due to COVID-19.

### 3.1.1 Sepsis and transcriptomic endotypes

The heterogeneity in sepsis has spurred in recent years a number of approaches to stratify patients beyond only looking at the microbial cause given the relevance of the host-pathogen interaction in any infectious disease. One such approach has been the use of peripheral blood transcriptomics to identify patterns of host response that may be specific to subgroups of patients. These groups of patients have been referred to as subphenotypes, endophenotypes or endotypes (Demerle et al. 2021; Shankar-Hari and Rubenfeld 2019; Stanski and Wong 2019), where subphenotypes refer to groups homogeneous for clinical and laboratory features (Reddy et al. 2020), endotypes groups with specific pathological mechanisms which may be amenable for therapeutic targeting, and endophenotypes a middle ground (Seymour et al. 2017). Below, I use the term endotype as the patient groups identified have distinct clinical outcomes and exhibit alterations of specific biological pathways suggesting true underlying disease mechanisms separating the groups, and the original authors have also termed the groups identified as endotypes.

Multiple groups have established a variety of definitions of transcriptomic endotypes in sepsis, with some of the earlier applications in pediatric patients (Wong et al. 2009; Wong et al. 2010; Wong et al. 2011; Wong et al. 2015; Wong et al. 2017b). In the definitions derived by Wong and colleagues, patients were stratified into endotype/subclasses A, B and C, with group A having more severe illness as measured by the pediatric risk of mortality III score (Pollack et al. 1997) and higher mortality (Wong et al. 2009). The 100 genes identified by the authors to accurately predict group membership were mostly downregulated in group A, and consisted of genes related to adaptive immunity and glucocorticoid signaling.

In adults, similar approaches have been deployed (Davenport et al. 2016;

Scicluna et al. 2017; Sweeney et al. 2018; Baghela et al. 2022). While Knight and colleagues (2016) separated patients into two endotypes, van der Poll and colleagues identified four endotypes (Scicluna et al. 2017), Sweeney and colleagues three (Sweeney et al. 2018) and Hancock and colleagues five (Baghela et al. 2022). All groups used variations of unsupervised clustering methods; for example, Knight and colleagues using agglomerative hierarchical clustering on the top 10% most variable genes. By contrast, van der Poll and colleagues utilized the top 5000 genes by mean absolute deviation, followed by consensus agglomerative hierarchical clustering on 1-Pearson correlation distances. Clustering stability was then evaluated through cumulative distribution functions, silhouette width analyses and cophenetic distance correlation analysis (Wilkerson and Hayes 2010; Rousseeuw 1987). The earlier studies (Davenport et al. 2016; Scicluna et al. 2017) were conducted in intensive care unit (ICU) patients, while the later ones included non-ICU patients (Sweeney et al. 2018; Baghela et al. 2022), possibly accounting for some of the differences in findings.

One stratification that has been shown to be reproducible across cohorts and that is robust to source of infection is the definition of sepsis responses signatures (SRS) established previously within the Knight group (Davenport et al. 2016; Burnham et al. 2017). Using microarray technology to probe the gene expression of total leukocytes from peripheral blood, two endotypes of sepsis patients were identified by unsupervised agglomerative hierarchical clustering based on the top 10% most variable probes ( $n = 2,619$ ) when initially applied to ICU sepsis patients suffering from community acquired pneumonia (CAP) (Davenport et al. 2016). The SRS1 signature displayed features of immunosuppression such as T cell exhaustion, HLA class II downregulation and endotoxin tolerance, potentially implicating T cells and monocytes. Importantly, SRS1 patients have been reproducibly found to have higher early mortality (14 or 28 day mortality

after sampling), and SRS was the predominant factor in separating patients in peripheral blood transcriptomic space independent of whether sepsis was caused by CAP or fecal peritonitis (FP) (Burnham et al. 2017). These SRS definitions could not be predicted by clinical covariates alone (Davenport et al. 2016), and association of SRS and 14-day mortality remained in a Cox proportional hazards model even after inclusion of age, sex and comorbidities as measured by the Charlson co-morbidity index (Burnham et al. 2017).

A previous attempt has been made to examine the overlap between pediatric and adult transcriptomic subphenotypes (Wong et al. 2017a). Pediatric endotype A/B definitions were applied to adult analysis, with a positive albeit weak correlation between SRS membership and pediatric endotype A/B definitions, and patients assigned jointly as pediatric endotype A and SRS1 had the highest mortality.

### **Therapeutic implications of transcriptomic subphenotypes**

Aside from prognostic implications, the endotypes identified have therapeutic consequences. For the pediatric stratifications, it was found that patients of endotype B had a substantially reduced risk of a complicated course of infection if given corticosteroids (Wong et al. 2017b), where the complicated course was defined as persistence until the seventh day of septic shock of  $\geq 2$  organ failures or mortality within 28 days.

In adults, a post-hoc analysis of the VANISH trial examining the effects of corticosteroids and vasopressin on patients admitted to ICU for septic shock demonstrated an interaction effect between corticosteroid usage and SRS group membership (Antcliffe et al. 2019). Patients receiving placebo recapitulated the differential survival signal by SRS (higher 14/28 day mortality in SRS1 patients), but the survival difference was no longer seen in the group given corticosteroids. This suggested that the immunocompetent SRS2 group may have in fact been harmed by the corticosteroid treatment.

### 3.1.2 COVID-19 peripheral blood bulk transcriptomics

While sepsis patients have previously been stratified on the basis of peripheral blood transcriptomics, understanding of COVID-19 risk has primarily been based on clinical parameters rather than stratification through molecular biomarkers or endotype definitions. Significant and common risk factors for a poor outcome in COVID-19 include age, male sex, obesity, diabetes and social deprivation (Williamson et al. 2020). This highlights the importance of considering and accounting for age, sex and comorbidities in any analysis of COVID-19 patients, as well as the remaining opportunities for using molecular biomarkers to stratify COVID-19 patients beyond using clinical features.

The peripheral blood transcriptome of COVID-19 patients was interrogated at the same time as the data for this chapter was generated (Thair et al. 2021; Aschenbrenner et al. 2021). Thair et al (2021) focused mostly on comparing the response to COVID-19 to that of other viral infections. Intriguingly, pathway analysis did not yield any statistically significant enrichments when comparing COVID-19 gene expression to non-COVID-19 respiratory (i.e. influenza, respiratory syncytial virus and SARS-CoV) viral gene expression. The analysis did not seek to stratify COVID-19 patients in the broader context of sepsis, which this chapter will discuss. Ascehnbrenner et al (2021) performed wider analysis to also include comparisons between COVID-19 and other viral/bacterial infections and other immune disorders. Further details are presented below in the discussion in context of the data presented here.

The data in this chapter have been released and published as part of the COMBAT consortium work (COMBAT consortium 2022), although the analyses presented below explore aspects of COVID-19 peripheral blood bulk transcriptomics in more detail and with additional perspectives including analysis in terms of SRS membership.

### 3.1.3 Aims

The overall aim of this chapter is to characterize the host peripheral blood transcriptomic response to sepsis in all-cause cases and in sepsis due to COVID-19. Specifically, I will look for shared and distinct aspects of the response to all-cause vs. COVID-19 specific sepsis. I aim to:

1. recapitulate SRS signatures with bulk RNA-seq in an all-cause and non-ICU specific sepsis cohort
2. confirm the existence and validity of SRS signatures in COVID-19 patients
3. characterize the differentially expressed genes and pathways between patients with all-cause sepsis and COVID-19 viral sepsis
4. identify the differences in gene expression in COVID-19 sepsis vs. non-septic COVID-19 illness

## 3.2 Results

### 3.2.1 Patient cohorts

In this chapter, I present the analysis of whole blood bulk RNA-seq data that I generated from healthy volunteers (HVs), patients hospitalized with mild, severe and critical COVID-19 infection (CM, CS and CC), patients in the community with mild COVID-19 infection (CComm), as well as all-cause sepsis patients (Sepsis). This dataset consisted of 144 samples (one filtered out from quality control, chapter 2) across n=124 individuals. COVID-19 cases were confirmed by nasopharyngeal swab positivity on clinical polymerase chain reaction (PCR) testing.

While all mild, severe and critical COVID-19 cases were hospitalized, they were categorized according to the level of respiratory or organ support required. Mild

COVID-19 denoted who required hospitalization but did not require any form of respiratory support. Critical COVID-19 referred to patients who required intensive care unit (ICU) admission for either intubation or developed circulatory shock and required vasopressor support. Severe COVID-19 included patients requiring any form of respiratory support lighter than invasive mechanical, ranging from supplemental oxygen to non-invasive positive airway pressure ventilation (bilevel positive airway pressure ventilation or continuous positive airway pressure ventilation) i.e. while not critical, nevertheless exhibited organ dysfunction associated with infection. The all-cause sepsis cohort spanned both severe (non-ICU) and critical (ICU) illness.

The clinical characteristics of the patient cohort are summarized in table 4.1.

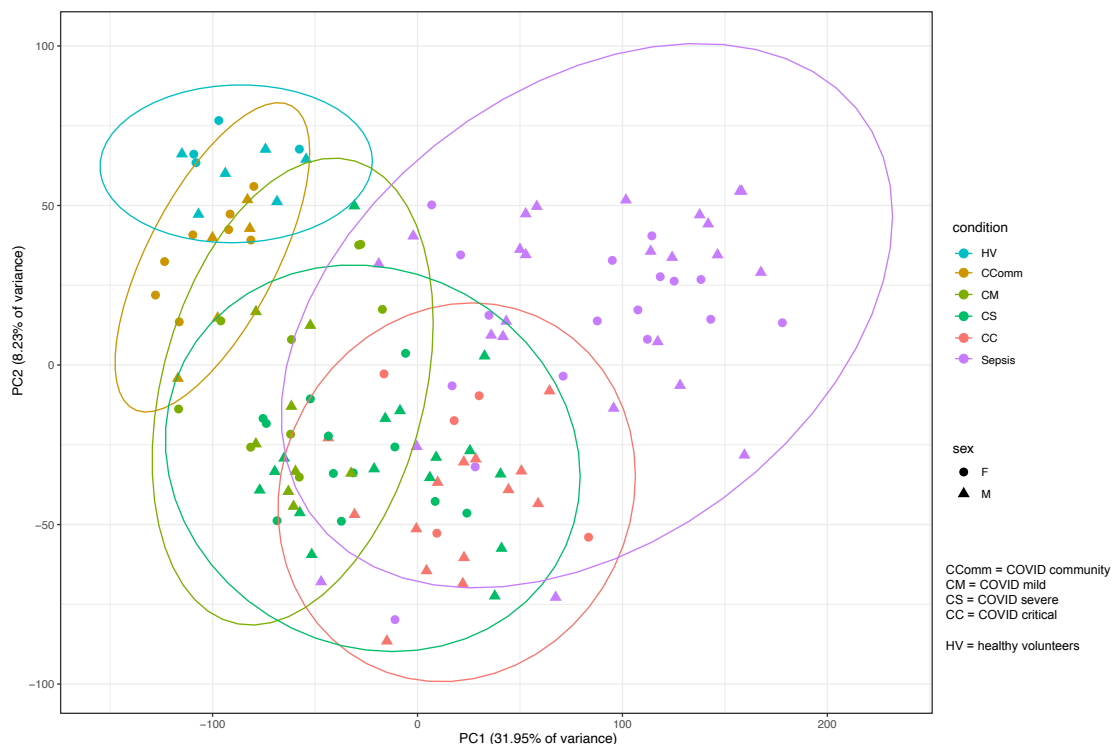
| Trait                                                         | HV (n=10)    | CComm (n=13) | CM (n=17)      | CS (n=28)      | CC (n=19)      | Sepsis (n=36)   | FDR     |
|---------------------------------------------------------------|--------------|--------------|----------------|----------------|----------------|-----------------|---------|
| Age, median (range)                                           | 60.5 (42-76) | 49 (34-59)   | 68 (21-94)     | 57.5 (22-78)   | 57 (37-78)     | 71 (20-92)      | 0.016   |
| Male sex, n (%)                                               | 6 (60%)      | 5 (38%)      | 8 (47%)        | 16 (57%)       | 14 (74%)       | 21 (58%)        | 0.45    |
| SOFA score, median (range)                                    | -            | -            | 0 (0-1)        | 2 (0-4)        | 7 (4-13)       | 3 (0-12)        | <0.0001 |
| SaO <sub>2</sub> /FiO <sub>2</sub> , median (range)           | -            | -            | 4.6 (4.4-4.8)  | 2.6 (1.0-4.0)  | 1.9 (1.1-2.7)  | 2.7 (0.9-4.7)   | <0.0001 |
| Charlson comorbidity index, median (range)                    | -            | -            | 0 (0-27)       | 4 (0-39)       | 0 (0-16)       | 0 (0-31)        | 0.27    |
| Estimated days from symptom onset to sampling, median (range) | -            | 14 (7-22)    | 6 (3-26)       | 11.5 (2-25)    | 13 (3-25)      | 5 (0-16)        | 0.01    |
| Length ICU stay, median (range)                               | -            | -            | 0 (0-0)        | 8 (2-21)       | 23 (8-56)      | 6 (0-37)        | 0.00014 |
| Ventilation days, median (range)                              | -            | -            | 0 (0-0)        | 0.5 (0-33)     | 21 (7-49)      | 1 (0-12)        | <0.0001 |
| Ever vasopressors, n (%)                                      | -            | -            | 0 (0%)         | 4 (14%)        | 17 (89%)       | 15 (42%)        | <0.0001 |
| Ever renal replacement therapy, n (%)                         | -            | -            | 0 (0%)         | 2 (7%)         | 6 (32%)        | 5 (14%)         | 0.035   |
| Death, n (%)                                                  | -            | 0 (0%)       | 0 (0%)         | 4 (14%)        | 3 (16%)        | 10 (28%)        | 0.1     |
| Laboratory findings, median (range)                           | -            | -            | -              | -              | -              | -               | -       |
| Leukocytes, x10 <sup>9</sup> /L                               | -            | -            | 7.0 (2.2-11.2) | 8.3 (3.8-15.2) | 8.5 (3.7-17.4) | 13.5 (1.5-40.9) | <0.0001 |
| Neutrophils, x10 <sup>9</sup> /L                              | -            | -            | 4.8 (1.1-8.3)  | 5.8 (1.7-13.4) | 7.5 (2.7-16.5) | 12.2 (1.2-38.4) | <0.0001 |
| Lymphocytes, x10 <sup>9</sup> /L                              | -            | -            | 1.3 (0.6-2.5)  | 1.0 (0.3-2.3)  | 0.7 (0.3-1.8)  | 0.9 (0.2-3.9)   | 0.027   |
| Monocytes, x10 <sup>9</sup> /L                                | -            | -            | 0.5 (0.1-1.4)  | 0.5 (0.2-1.1)  | 0.4 (0.0-1.0)  | 0.6 (0.0-2.5)   | 0.035   |

**Table 3.1: Clinical characteristics of bulk RNA-seq cohort of n=124 individuals.** Kruskal-Wallis tests performed for continuous data and chi-square test performed for count data. Data are n (%) or mean (SD) unless otherwise specified.



### 3.2.2 Unsupervised analysis shows separation of disease states

I first performed unsupervised analysis of the whole dataset with PCA and hierarchical clustering on all 23,063 genes. PCA decomposes the variance structure in the dataset to extract orthogonal axes of variation in order of variance accounted for. PCs 1 and 2 accounted for 31.95% and 8.23% of total variance respectively. The six comparator groups HVs, CM, CS, CC, CC and sepsis were separated clearly across the first two PCs despite no selection of highly variable genes, with PC1 separating the samples more by severity of disease overall (HV had the lowest PC1 values and CC/sepsis had the highest PC1 values), while PC2 also separated COVID samples by severity, although CS and CC samples had similar PC2 scores. This suggested that the peripheral blood transcriptome was reflective of clinical groupings of the patients in this cohort and that there was distinct biology underlying the clinically observed phenotypes.

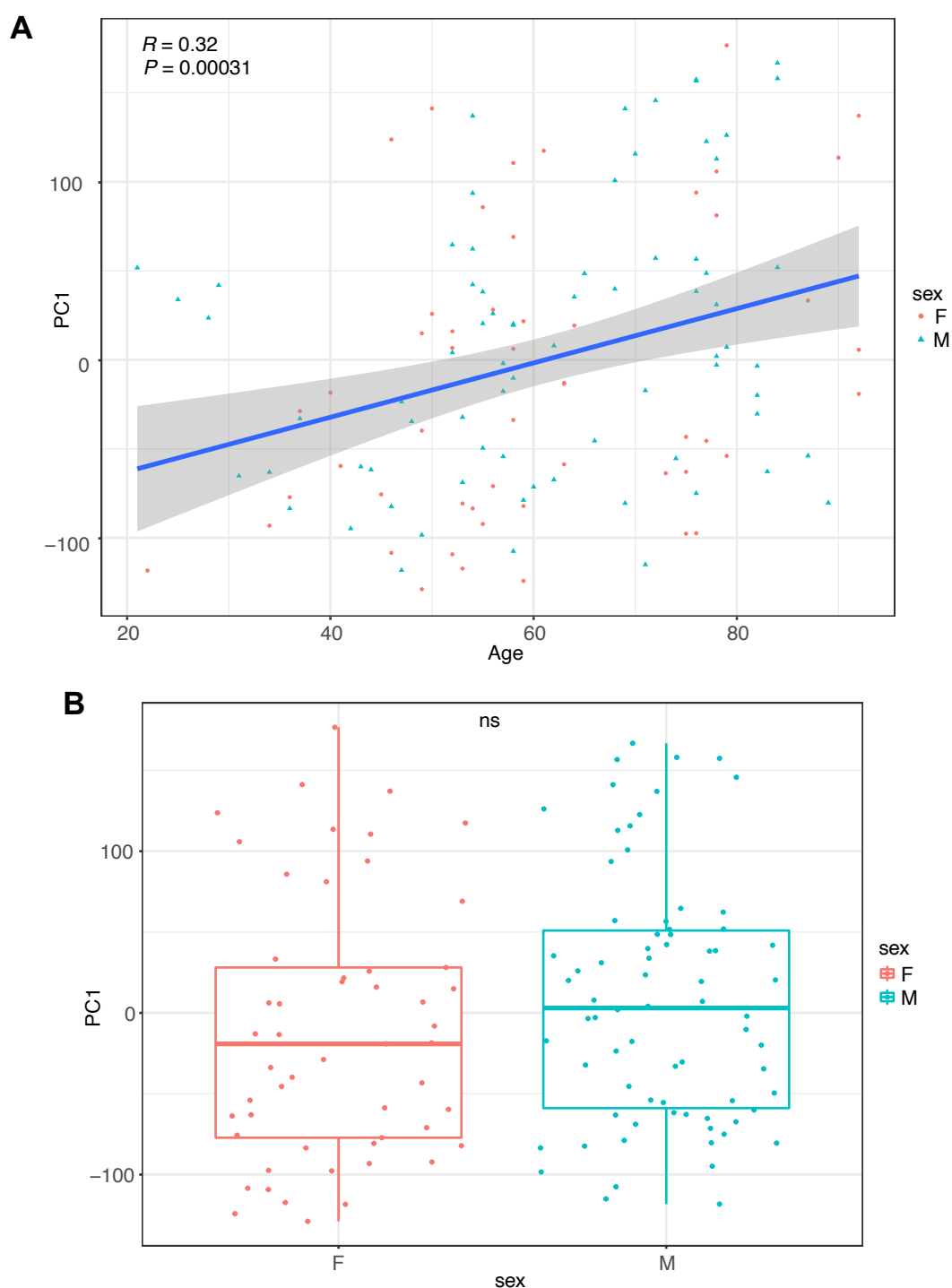


**Figure 3.1: Principal components analysis (PCA) of bulk RNA-seq dataset.** PC1 and 2 of PCA on all 23,063 genes on n=143 samples.

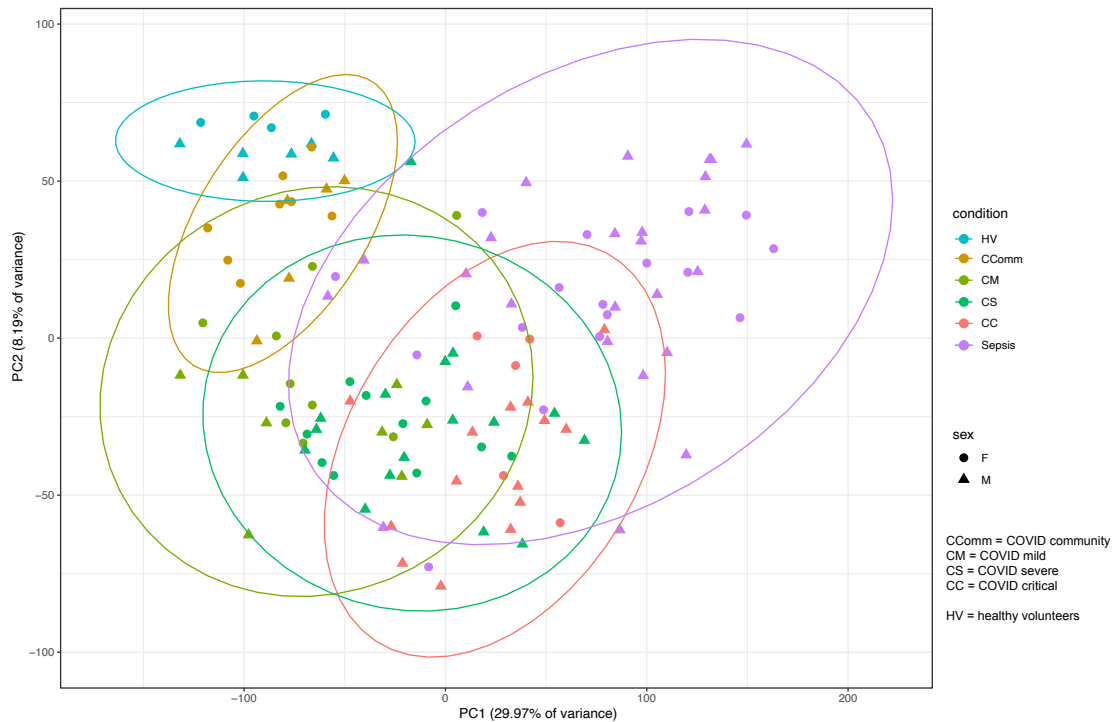
Given the association of severe disease with older age and male sex in COVID-19, I examined the difference in PC1 score across the cohort by sex and tested for correlation of age with PC1 score. A statistically significant correlation between PC1 and age was found ( $P < 0.001$ ), but with a weak correlation of only  $r=0.32$  (Fig. 3.2A). There was no significant difference between sex for PC1 score (Fig. 3.2B). To confirm the minimal effects of age and sex, I fitted a linear model for sex, age and aged squared terms for each gene on the normalized and filtered count matrix before repeating PCA, after which it could be seen that the separation in various comparator groups remained clear and in an almost identical pattern, and the percentages of variance accounted for by the first two PCs were only minimally different (29.97% and 8.19% for PCs 1 and 2 respectively) (Fig. 3.3).

I then used a different method of unsupervised analysis, namely agglomerative hierarchical clustering. I found that the grouping of samples was broadly concordant to PCA i.e. grouping mostly by comparator group, regardless of whether clustering with Ward's algorithm was performed on all genes based on Euclidean distance (Fig. 3.4A) or based on sample correlations (Fig. 3.4B). Cutting the clustering tree into three clusters in either instance, both methods identified a cluster of only non-COVID-19 all-cause sepsis, and also split the other samples with a cluster containing HVs and CComm as well as CM and CS samples but no CC samples, and another cluster containing the majority of CC samples and some sepsis samples, but none of the HV and CComm samples.

Taken together, the unsupervised analyses suggested that peripheral blood bulk transcriptomics reflect clinical severity and therefore represents a meaningful way to understand potential pathogenic mechanisms underlying COVID-19 and sepsis.



**Figure 3.2: Association of PC scores with age and sex.** (A-B) PC1 (C-D) PC2. (A) Correlation of PC1 score with age for each sample and (B) PC1 score distributions by sex. (C-D) Correlation and distributions of PC2 scores.

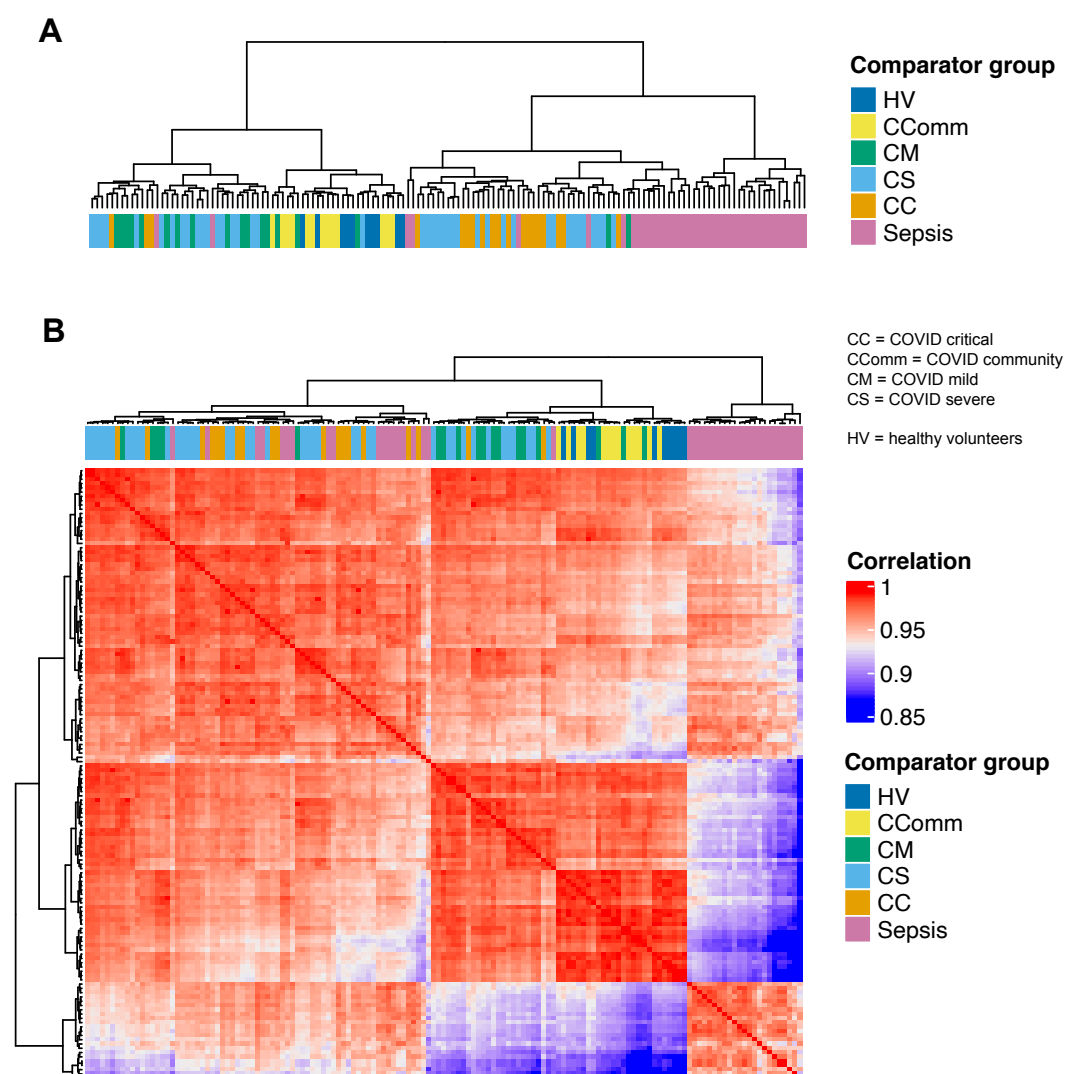


**Figure 3.3: Principal components analysis (PCA) of bulk RNA-seq dataset accounting for age and sex.** PC1 and 2 of PCA on all 23,063 genes on n=143 samples after fitting a linear model per gene regressing out age and sex.

### 3.2.3 Recapitulation of SRS signatures in all-cause sepsis cohort

Previous work had defined disease endotypes in sepsis that spanned sepsis of different sources and infectious aetiologies (Burnham et al. 2017; Antcliffe et al. 2019). However, the endotypes were derived strictly from patients admitted to the ICU. To investigate whether a similar endotype definition could be recapitulated in the non-COVID all-cause sepsis cohort that included patients not admitted to the ICU, I performed separate unsupervised analysis on only the sepsis samples.

The previous definitions of SRS were established by either agglomerative hierarchical clustering, or endotype assignment with a seven gene set (*DYRK2*, *CCNB1IP1*, *TDRD9*, *ZAP70*, *ARL14EP*, *MDC1* and *ADGRE3*). I applied this seven gene set definition to the new samples for SRS assignment. This was done by



**Figure 3.4: Hierarchical clustering analysis.** (A) Agglomerative hierarchical clustering with Ward's method on euclidean distance of all 23,063 genes for n=143 samples. (B) Agglomerative hierarchical clustering based on sample sample correlational distances with 23,063 genes and corresponding correlation heatmap.

utilizing a logistic regression model trained on expression of the seven genes on an in house dataset of sepsis samples from community acquired pneumonia (CAP) and fecal peritonitis (FP) patients to assign SRS for the current all-cause sepsis cohort. Taking the most recent time point available, this yielded 26 SRS 1 samples (21 individuals) and 16 SRS 2 samples (15 individuals).

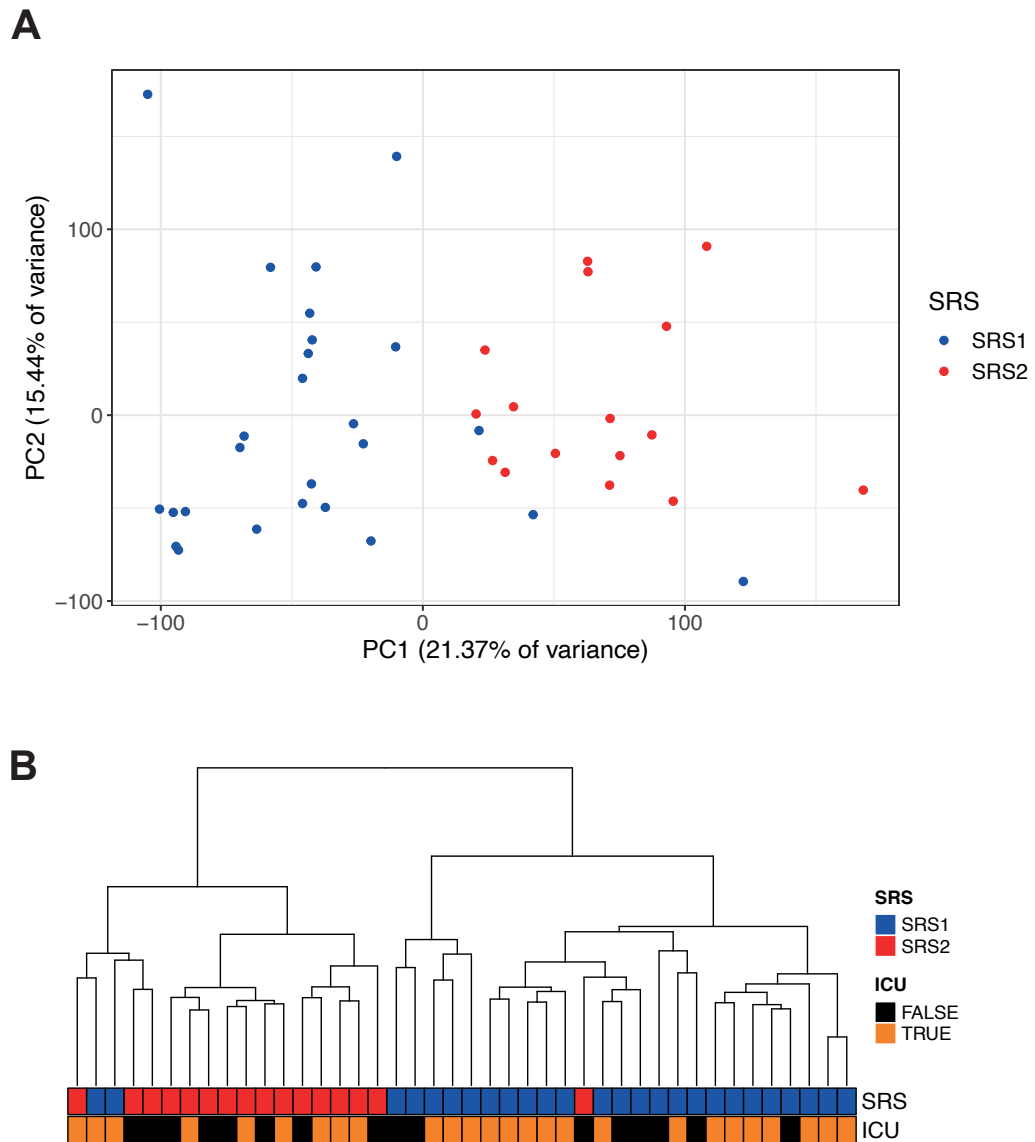
PCA on all genes of these samples showed that samples defined as SRS 1 and SRS 2 were separated along PC1 which accounted for 21.37% of total variance (Fig. 3.5A). Meanwhile, agglomerative hierarchical clustering on all genes for these samples showed that at the level of two clusters, the samples were split mainly by SRS with only 3 samples clustering discordantly to their SRS definition (Fig. 3.5B), while ICU status on sampling was instead relatively evenly split across the two clusters.

The unsupervised analysis suggested that the original SRS definition based on microarray data and from ICU cohorts could be faithfully recapitulated in an all-cause sepsis cohort including non-ICU patients.

### **3.2.4 Validation of SRS signatures with consensus SRS definition**

As previous SRS definitions were established specifically with peripheral blood microarray based gene expression experiments, I utilized a newly derived consensus SRS endotyping technique (Cano-Gamez et al. 2022) to further interrogate the stability of SRS assignment in these samples with the Sepstratifier package.

In brief, the Sepstratifier package takes the expression of the SRS seven gene set and aligns the samples of interest to a set of gene expression profiles collated from sepsis patients and healthy individuals via a mutual nearest neighbours (MNN) algorithm. As this reference was built containing gene expression data



**Figure 3.5: Unsupervised exploratory data analysis of non-COVID-19 sepsis samples.** (A) Principal components analysis (PCA) on all 23,063 genes for  $n=42$  sepsis samples colored by sepsis response signature (SRS) assignment. (B) Agglomerative hierarchical clustering with Ward's method on euclidean distance of all 23,063 genes for  $n=42$  sepsis samples with SRS assignment and intensive unit care status shown

| SRS old \ SRS new | SRS1 | SRS2 | SRS3 |
|-------------------|------|------|------|
| SRS1              | 21   | 0    | 0    |
| SRS2              | 0    | 15   | 1    |

**Table 3.2: Sepsis sample sepsis response signature (SRS) assignments.** Table showing overlaps in SRS assignments by random forest model implemented in SepstratifieR (columns) or logistic regression model from training on in-house sepsis RNA-seq data as part of the GAINs study (rows).

from qPCR, microarray and bulk RNA-seq platforms, it mitigates the issue of differences in technical platform in the current bulk RNA-seq dataset vs. the original definitions of SRS via microarray. The MNN step aligns the reference and query samples to the same scale. The definition of SRS is then performed with a random forest model trained on the reference samples.

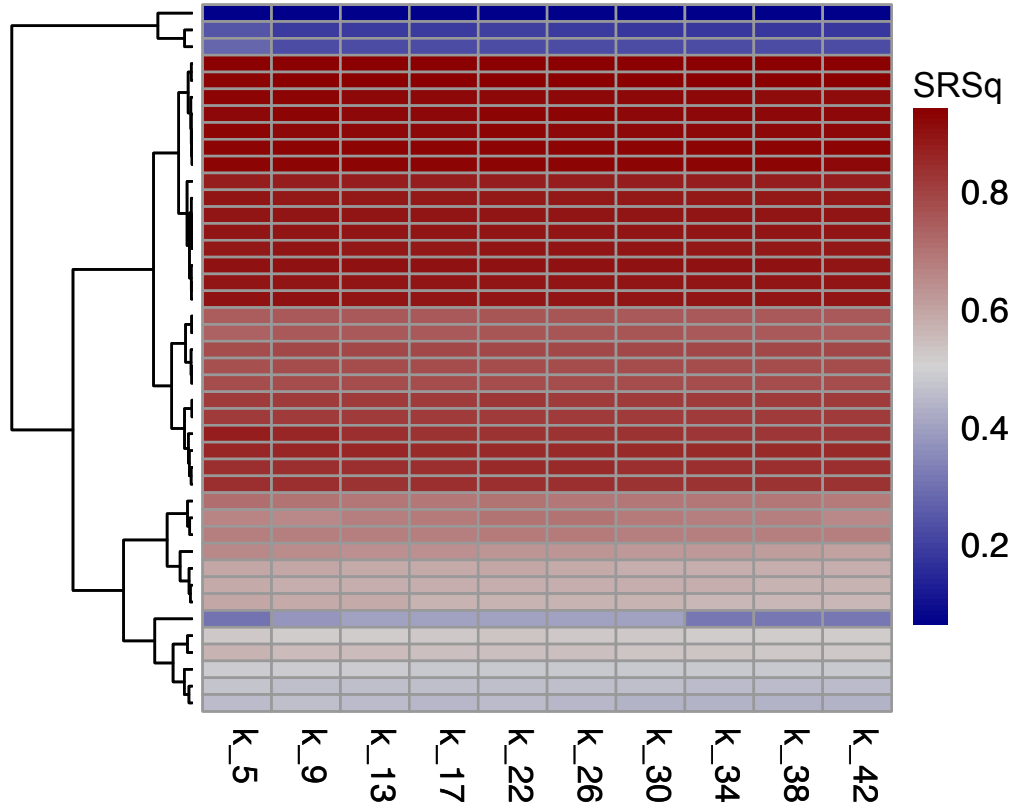
Importantly, one extension of this method over the original derivation of SRS is the introduction of SRS3 as a separate classification, with the group consisting of mostly healthy individuals in the reference dataset. This therefore expanded beyond the original definition of SRS as a metric to resolve heterogeneity within sepsis, to a metric considering a spectrum from healthy to sepsis with SRS1 indicating a group of patients with a molecular signature associated with particularly severe sepsis. Differential testing across SRS state from here onwards therefore refer to comparisons of SRS1 vs. non-SRS1 (i.e. SRS2 + SRS3) samples.

Using this method, I found that there was only one difference in SRS classification with the day 1 sample from patient N00006 (N00006.Ja001T.TRGa) classified as SRS3 (and not SRS2 as in the original classification) (Table 3.2). All other samples retained the same SRS1 / SRS2 classification, verifying the robustness of both original and new definitions of SRS.

A second key feature of the SepstratifieR package in addition to the novel SRS3 class is the derivation of a quantitative metric of SRS status (SRSq score). The motivation behind such a score is that the severity of patients' illness often span a continuous spectrum. A second random forest model is therefore used to output



the SRSq score between 0 to 1, which is based on the score of the first diffusion component of a diffusion map dimensionality reduction of the expression of the seven SRS genes.



**Figure 3.6: Sensitivity analysis for Sepstratifier sepsis response signature (SRS) assignment.** The  $k$  hyperparameter in SRS assignment by Sepstratifier (number of nearest neighbors to use for aligning query to reference data) has the greatest potential in altering assignment results. The figure shows a heatmap of SRSq scores per sample across a range of values for  $k$ .

To evaluate the stability of the outputs from Sepstratifier, I performed a sensitivity analysis across different choices of the hyperparameter  $k$  (the number of mutual nearest neighbours in the alignment step between reference and query samples) (Fig. 3.6). This hyperparameter is the most important one to consider as a high value results in stronger alignment even if query and reference data are substantially different, and vice versa. A high value therefore may improve data integration but forcibly align outliers which should not be aligned. Varying  $k$  from a range between 5 to 42, the predicted SRSq scores remained similar

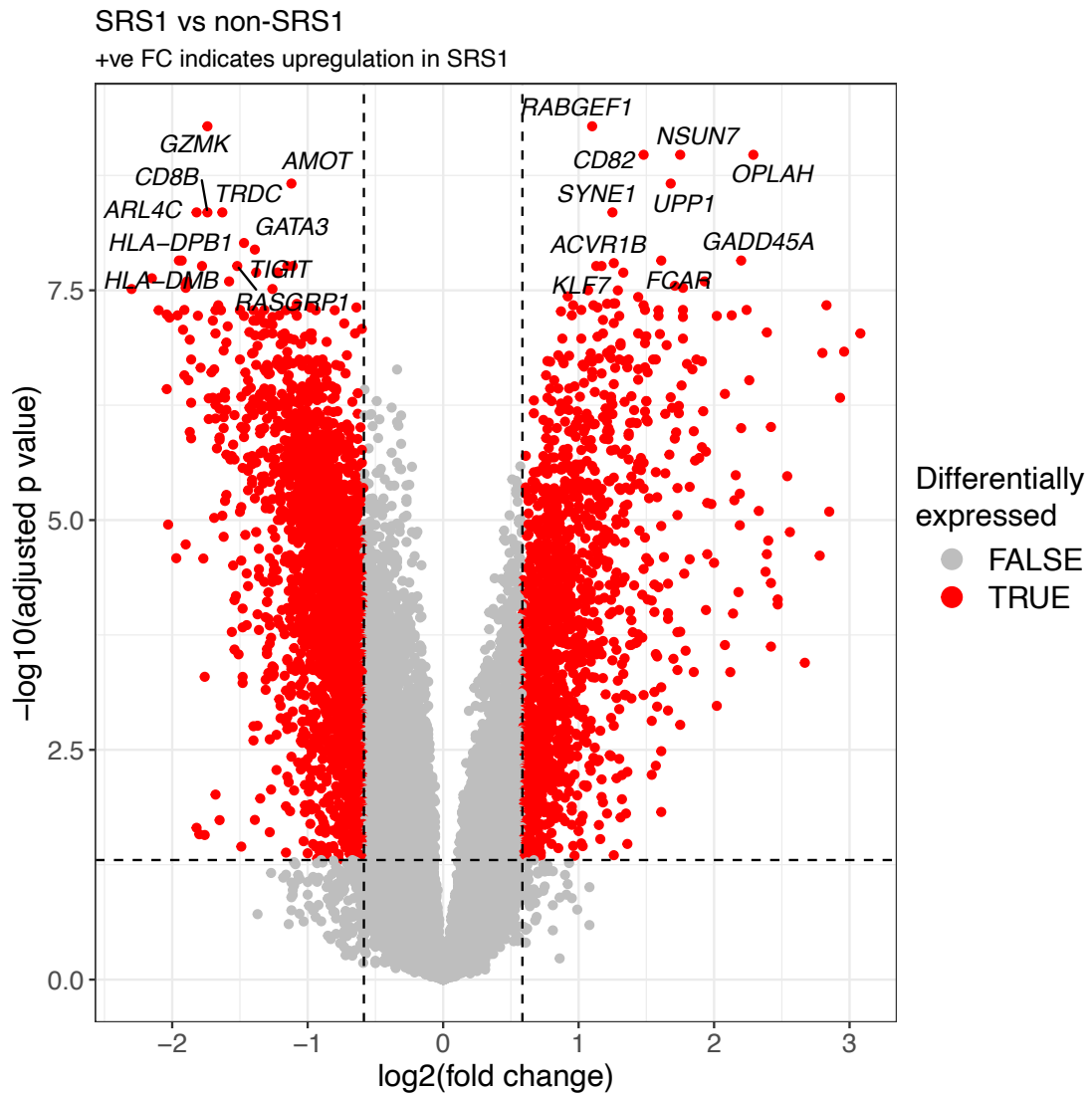
throughout, suggesting that the SRSq scoring and SRS classification was robust.

### **3.2.5 Validation of SRS assignment with differential gene expression and mortality analysis across SRS for all-cause sepsis**

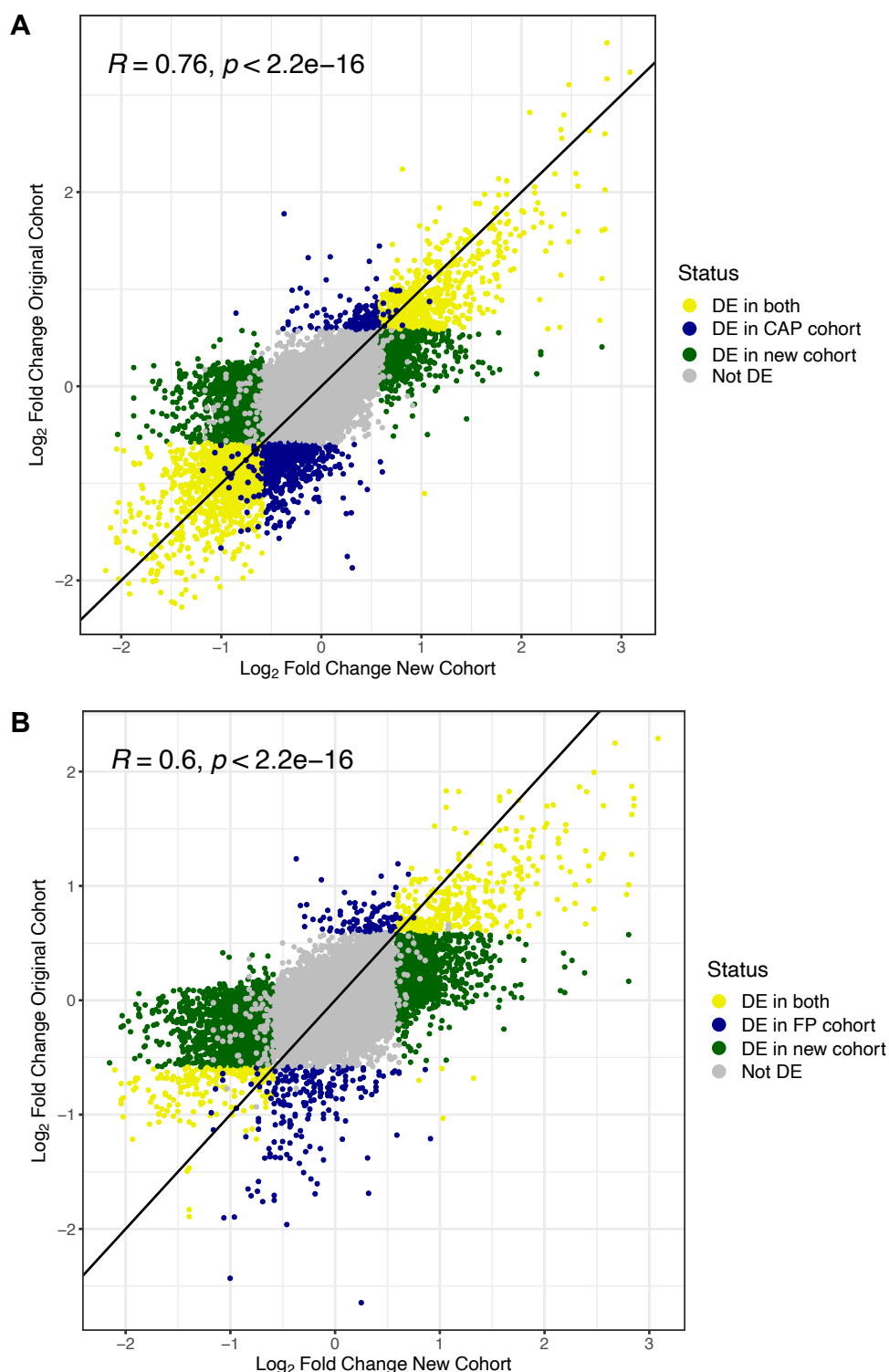
Having established the assignment of SRS endotypes in the non-COVID all-cause sepsis cohort, I then searched for further evidence that the algorithmic assignments were interpretable in the context of the original SRS definitions. To this end, I undertook mortality analysis and differential gene expression (DGE) analysis of the all-cause sepsis samples. I first performed DGE analysis on SRS1 vs. non SRS1 samples with a linear mixed model to account for repeated sampling of six individuals (n=42 samples total). This yielded 4030 differentially expressed genes at thresholds of fold change (FC)  $\geq 1.5$  and false discovery rate (FDR) of  $< 0.05$ .

I then intersected the genes identified in this experiment with the genes identified in the experiments where SRS was originally defined in the CAP and FP cohorts, and compared the log fold change (LFC) of these genes across the cohorts. I found 10362 commonly identified genes in comparing the current dataset with the original CAP cohort, and 10446 genes between the current dataset and the FP cohort. As the original SRS comparisons were made using SRS as a categorical variable, I examined the correlation of the LFC between genes in SRS1 vs. SRS2 in the original CAP cohort and in SRS1 vs. non-SRS1 samples in the current cohort, finding a significant correlation with a Pearson's  $r$  of 0.76 and  $P < 2.2e-16$ . Similarly, there was a correlation of  $r = 0.6$  and  $P < 2.2e-16$  when comparing LFC between genes in SRS1 vs. SRS2 in the FP cohort against the current sepsis cohort.

Another feature of the original SRS groupings was that SRS1 individuals

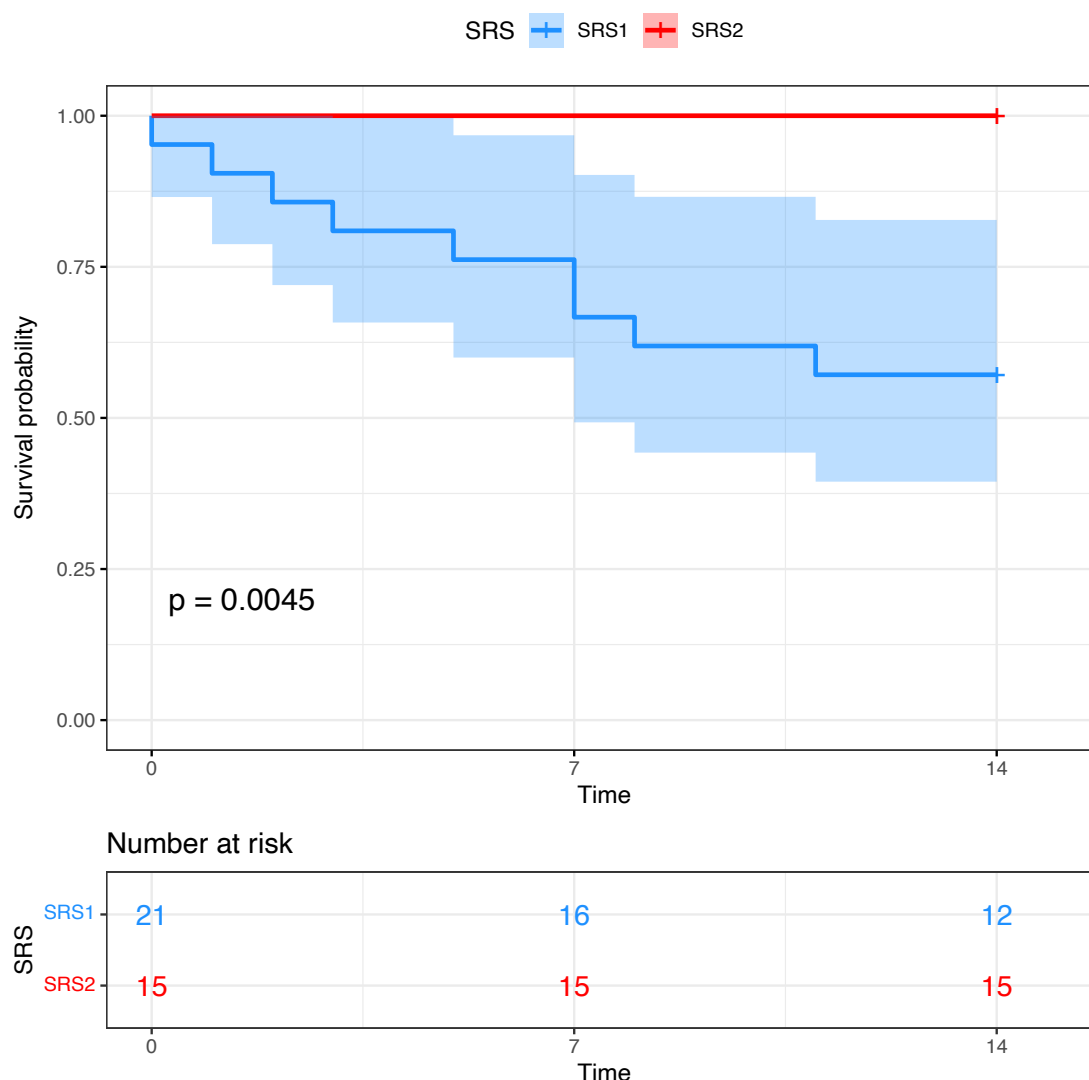


**Figure 3.7: Differential gene expression (DGE) across sepsis response signature groups.** Volcano plot of DGE for SRS1 (n=26 samples) vs. SRS2 (n=16 samples) patients. FC = fold change, Differentially expressed = FC  $\geq 1.5$  and adjusted p-value  $< 0.05$ .



**Figure 3.8: Assessing validity of sepsis response signature (SRS) assignment by correlation with original cohort SRS differential expression.** (A) Correlation of log fold change between SRS1 vs. SRS2/non-SRS1 in current sepsis cohort against original cohort from which SRS was derived (community acquired pneumonia [CAP] patients) (Davenport et al. 2016). (B) Analysis as per (A) with current cohort against fecal peritonitis (FP) patients from Burnham et al (2017).

exhibited higher early mortality than SRS2/non-SRS1 patients. Therefore, beyond differential gene expression correlations, I examined whether a similar pattern or higher early mortality in SRS1 patients could be found. Indeed, Kaplan-Meier curve survival analysis showed that there was higher 14 day mortality in SRS1 patients vs. non-SRS1 patients (log rank test  $p=0.0045$ ) (Fig. 3.9). A Cox's proportional hazard could not however be run due to the lack of deaths in the non-SRS1 group, which would result in a degenerate estimate in the group with no events.



**Figure 3.9: Mortality analysis by sepsis response signature (SRS).** Kaplan-Meier survival curve (top) for time after sampling stratified by SRS with corresponding risk table (bottom).  $P$ -value calculated by log-rank test.

|        | SRS1 | SRS2 | SRS3 |
|--------|------|------|------|
| Number | 3    | 49   | 39   |

**Table 3.3: COVID-19 sample sepsis response signature (SRS) assignments.** Table showing SRS assignments by random forest model implemented in SepstratifierR for COVID-19 RNA-seq samples.

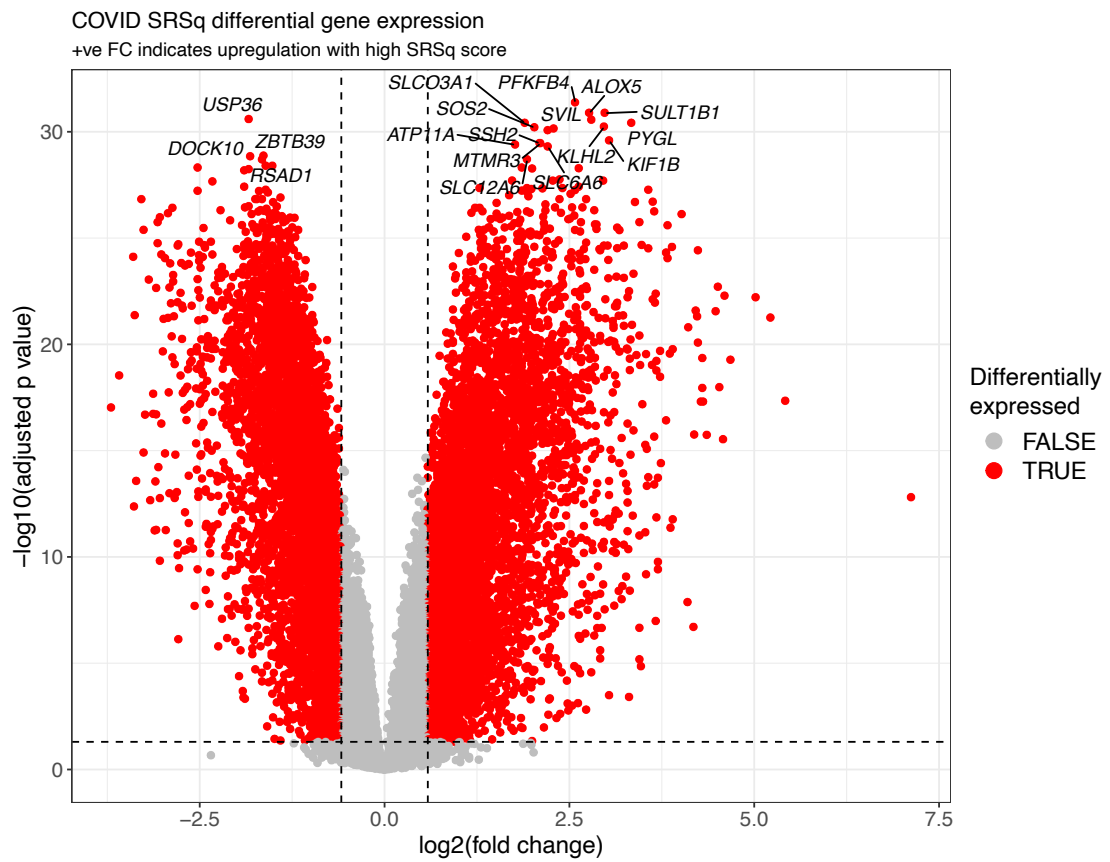
### 3.2.6 Application of SRS to COVID-19 samples

Having established the suitability of using SRS as a framework to analyse the whole blood bulk RNA-seq data generated from all-causes sepsis inclusive of both ICU and non-ICU patients, I next investigated the applicability of SRS to patients with COVID-19. Again using the framework from the SepstratifierR package, I classified the various COVID-19 samples into 3 SRS groups.

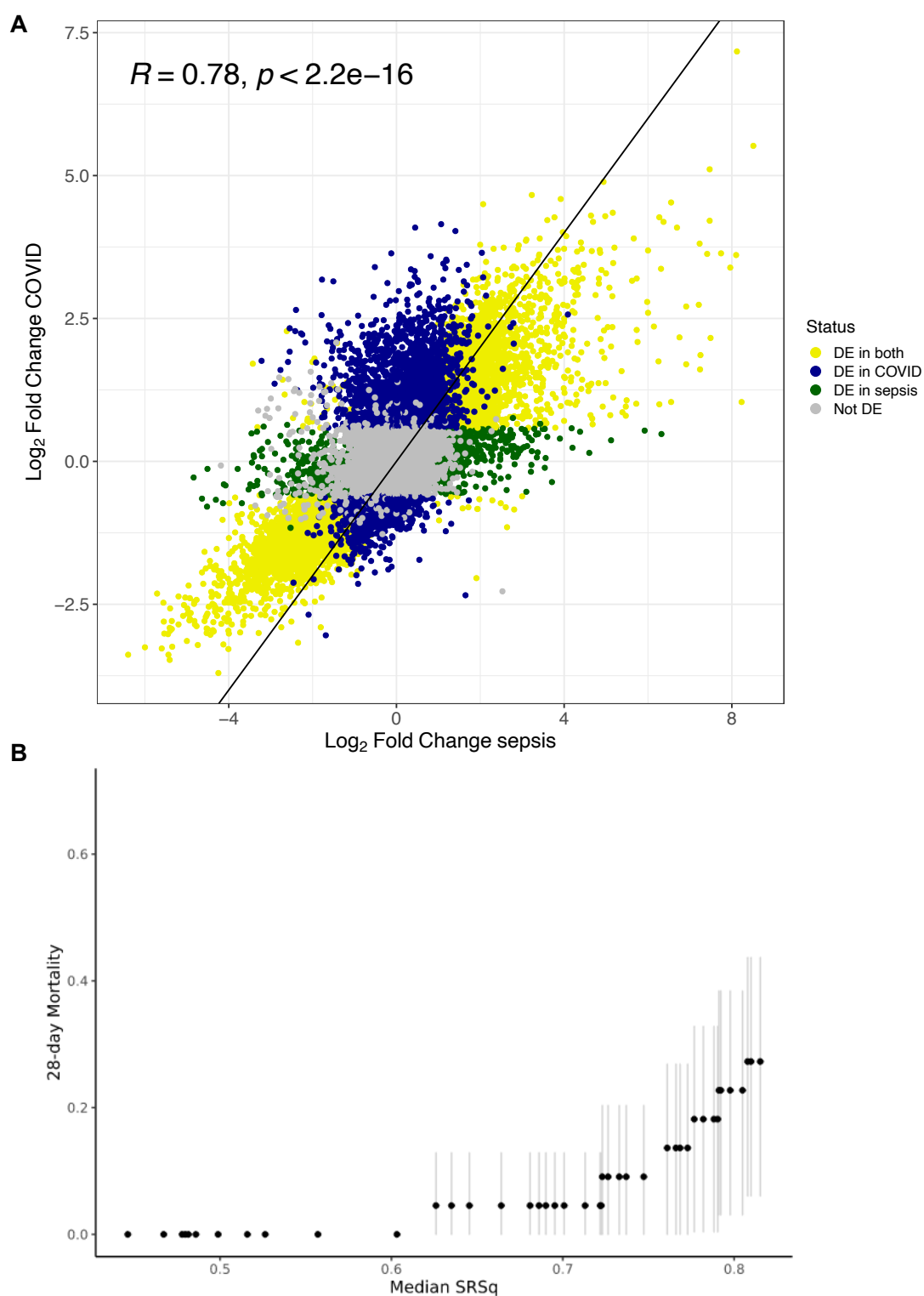
With this analysis, I found that there were only 3 SRS1 samples, with the majority of the samples labelled SRS2 (49) and one-third SRS3 (39) (Table 3.3). Given the low number of SRS1 samples, a DGE analysis of SRS as a categorical trait would be underpowered and as such I tested for DGE based on SRS as a quantitative trait using the SRSq score.

Without accounting for any other covariates, DGE revealed 10448 differentially expressed genes (Fig. 3.10). I compared these genes to the DEGs from all-cause sepsis SRS analysis by correlating the log fold changes by SRS which showed a strong and highly significant correlation (Pearson's  $r = 0.78$ ,  $P < 2.2e-16$ ) (Fig. 3.11A).

Similar to the SRS-mortality analysis in the all-cause sepsis cohort, I sought evidence for a relationship between mortality and SRS status in the COVID-19 samples. However, given that there were only 3 patients assigned an SRS1 status, I examined the rolling average of 28 day mortality with SRSq score. This revealed an increase in 28-day mortality associated with higher SRSq scores ( $P = 0.0028$ , Cox proportional hazards model) (Fig. 3.11B), providing further evidence of the validity of SRS in the setting of COVID-19. Together, these results suggested that



**Figure 3.10: Differential gene expression by SRSq for COVID-19 samples.** Volcano plot of SRS DGE for COVID-19 samples (n=91) using SRS as a quantitative trait (SRSq scores). FC = fold change, Differentially expressed =  $\text{FC} \geq 1.5$  and adjusted p-value  $< 0.05$



**Figure 3.11: Assessing validity of sepsis response signature (SRS) assignment in COVID-19 samples.** (A) Correlation of gene expression log fold change across SRSq in COVID-19 samples vs. sepsis samples. Mortality analysis plot of 28-day mortality vs. SRSq score with 95% confidence intervals shown.



the extended SRS framework was a relevant when applied to COVID-19 samples as well.

### **3.2.7 Differential gene expression (DGE) and pathway analysis**

To understand the biology behind the immune response in sepsis, SRS signatures, and the shared versus distinct aspects between all-cause vs. COVID-19 induced viral sepsis, I conducted DGE followed by pathway enrichment analysis across a number of comparisons.

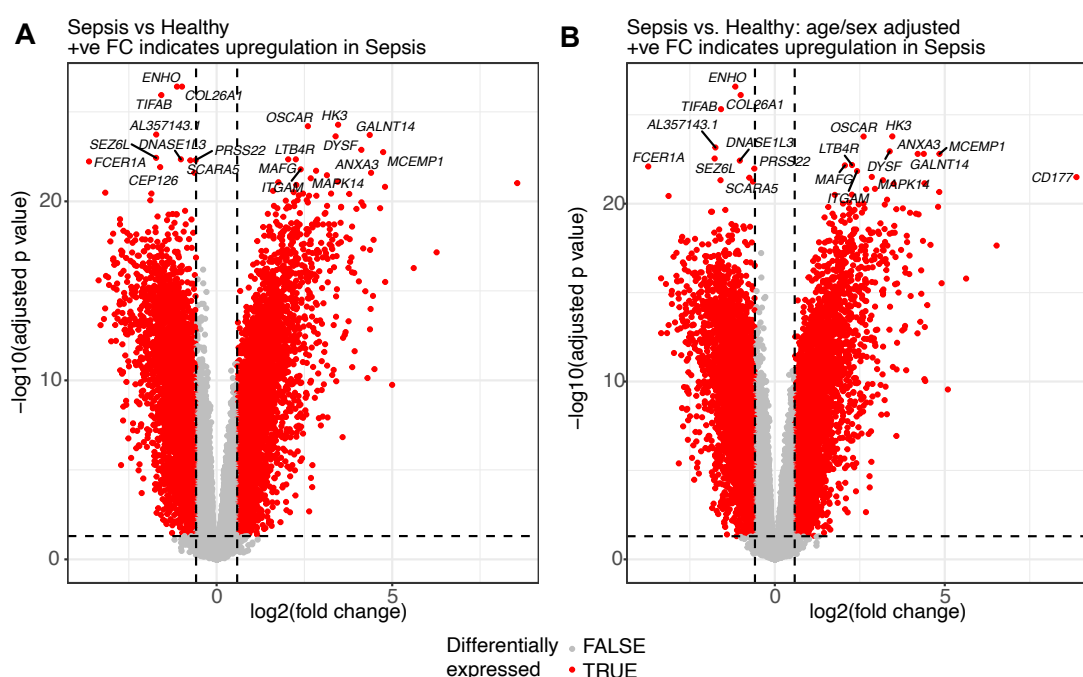
I performed DGE with a linear mixed model on normalized data for:

1. All-cause sepsis vs. healthy volunteers. This was to first characterize the overall transcriptomic response in sepsis.
2. Within all cause sepsis: SRS1 vs. non SRS1. This was to test for whether the transcriptomic characteristics of SRS could be recapitulated with RNA-seq and in a cohort including non-ICU patients.
3. All cause sepsis vs. severe and critical COVID-19 patients. This was to test the hypothesis that there are distinct transcriptomic aspects of COVID-19 sepsis.
4. COVID-19 critical patients vs. COVID-19 mild patients. This was an alternative viewpoint to understand the specific transcriptomic signatures to COVID-19 sepsis.

#### **The sepsis transcriptomic response**

I first dissected the transcriptomic response to sepsis by performing DGE analysis between sepsis (n=42 samples) and HVs (n=10 samples). This yielded 8751 genes differentially expressed at  $FC \geq 1.5$  and  $FDR < 0.05$  (Fig. 3.12A). Accounting for age and sex in the linear model changed the number of DEGs to 8629, suggesting minimal contributions from these covariates, as per the PC analysis earlier (Fig.

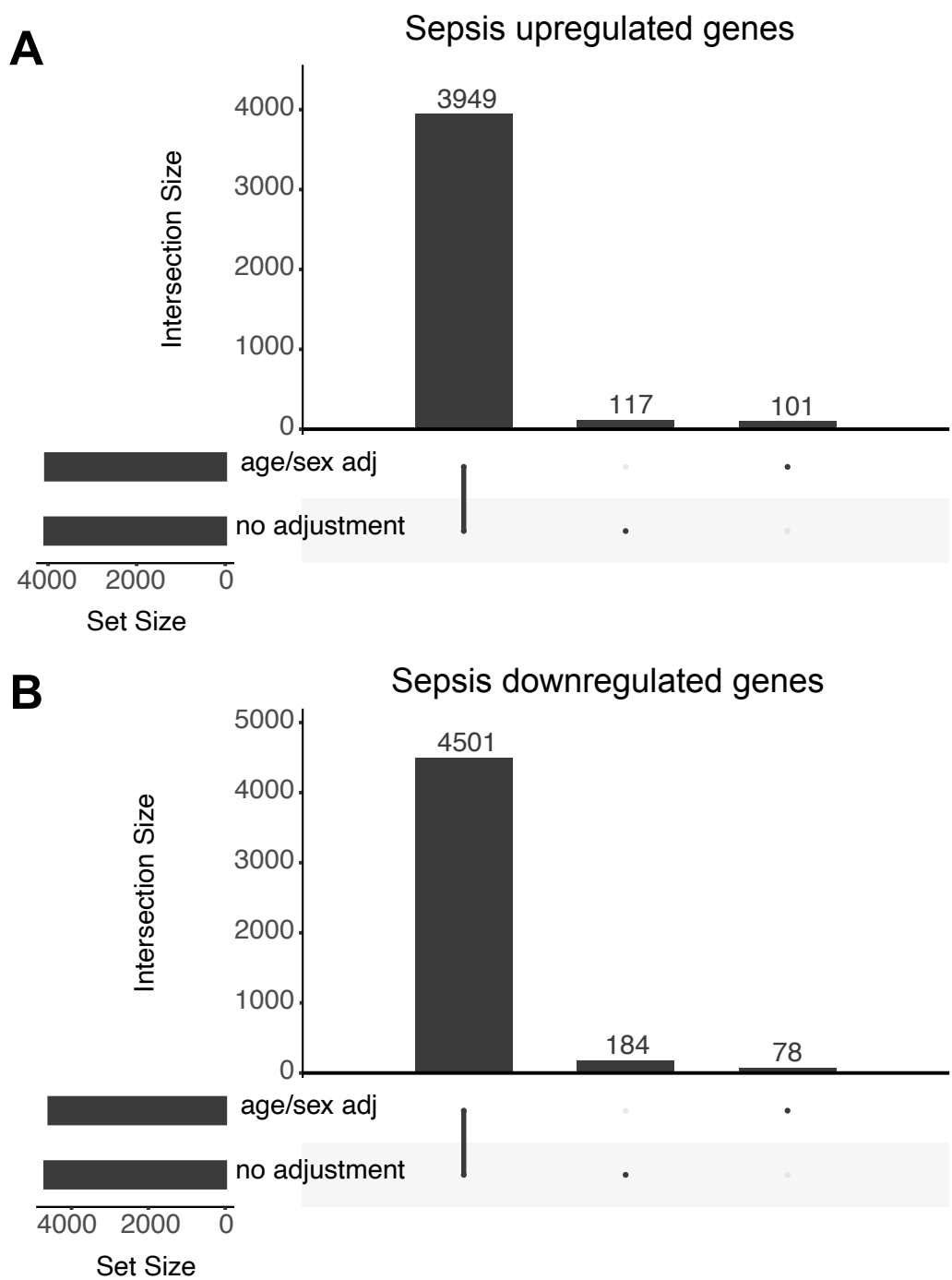
3.12B). On overlapping either the up or downregulated genes, the vast majority were the same with or without covariate inclusion into the model (Fig. 3.13A-B). Pathway analyses with the Gene Ontology Biological Processes (GOBP) database showed enrichment of immune pathways in both up and downregulated genes. NF- $\kappa$ B signaling as well as general immune system pathways were enriched in sepsis, while there was enrichment for reduced T cell activation and the humoral immune response in sepsis.



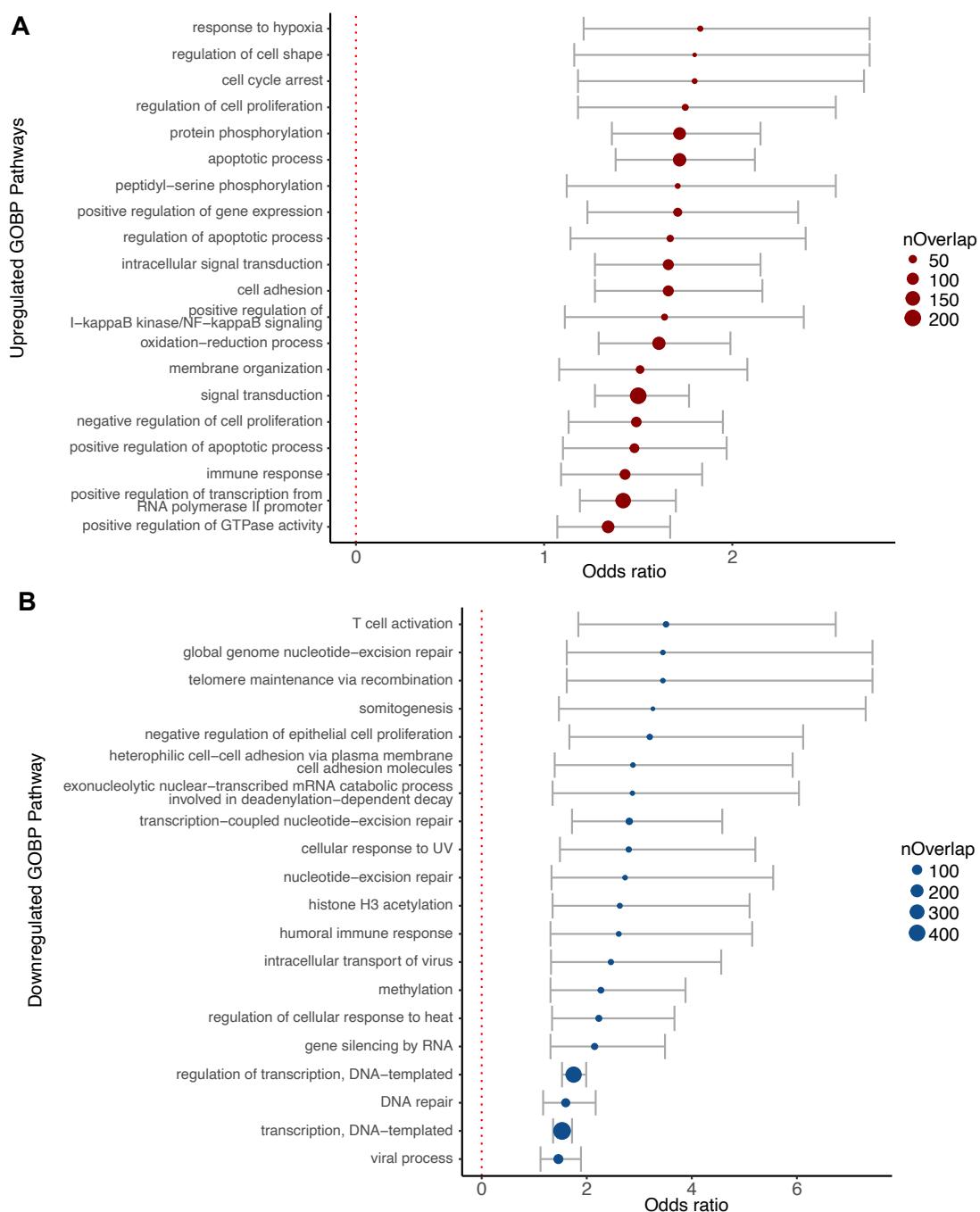
**Figure 3.12: Differential gene expression (DGE) of sepsis vs. healthy volunteer (HV) samples.** (A-B) Volcano plots for DGE of sepsis (n=42) vs. HV (n=10) samples, (A) without accounting for any covariates and (B) including age and sex into the DGE model. FC = fold change, Differentially expressed = FC  $\geq 1.5$  and adjusted p-value  $< 0.05$

### Transcriptomic differences across SRS

To further understand the heterogeneity within the sepsis response, I examined the gene expression changes across SRS1 (n=26) vs. non-SRS1 (n=16) samples. Testing SRS as a categorical trait resulted in 4030 DEGs (Fig. 3.7). Accounting for age and sex in the linear model marginally increased the number of DEGs to 4477 genes (Fig. 3.15A). To understand whether the differential expression may have been due to gross changes in leukocyte populations, I utilized the clinical

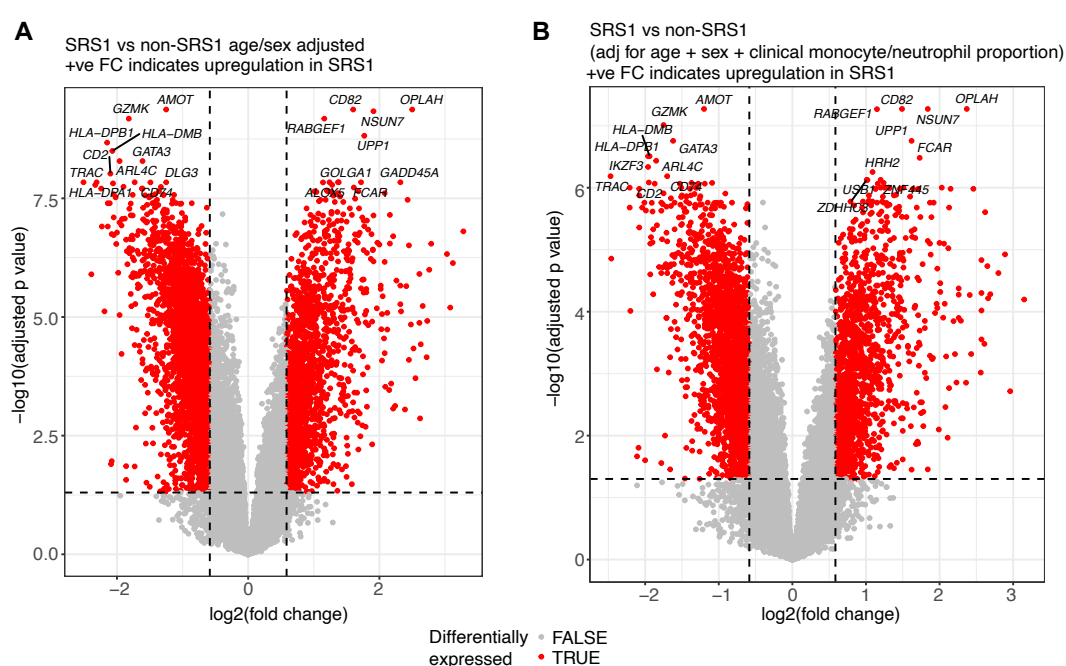


**Figure 3.13: Differential gene expression (DGE) overlaps of sepsis vs. healthy volunteer (HV) samples with and without covariates.** (A-B) Overlaps in (A) upregulated and (B) downregulated genes when comparing sepsis vs. HV samples including vs. not including age and sex covariates.



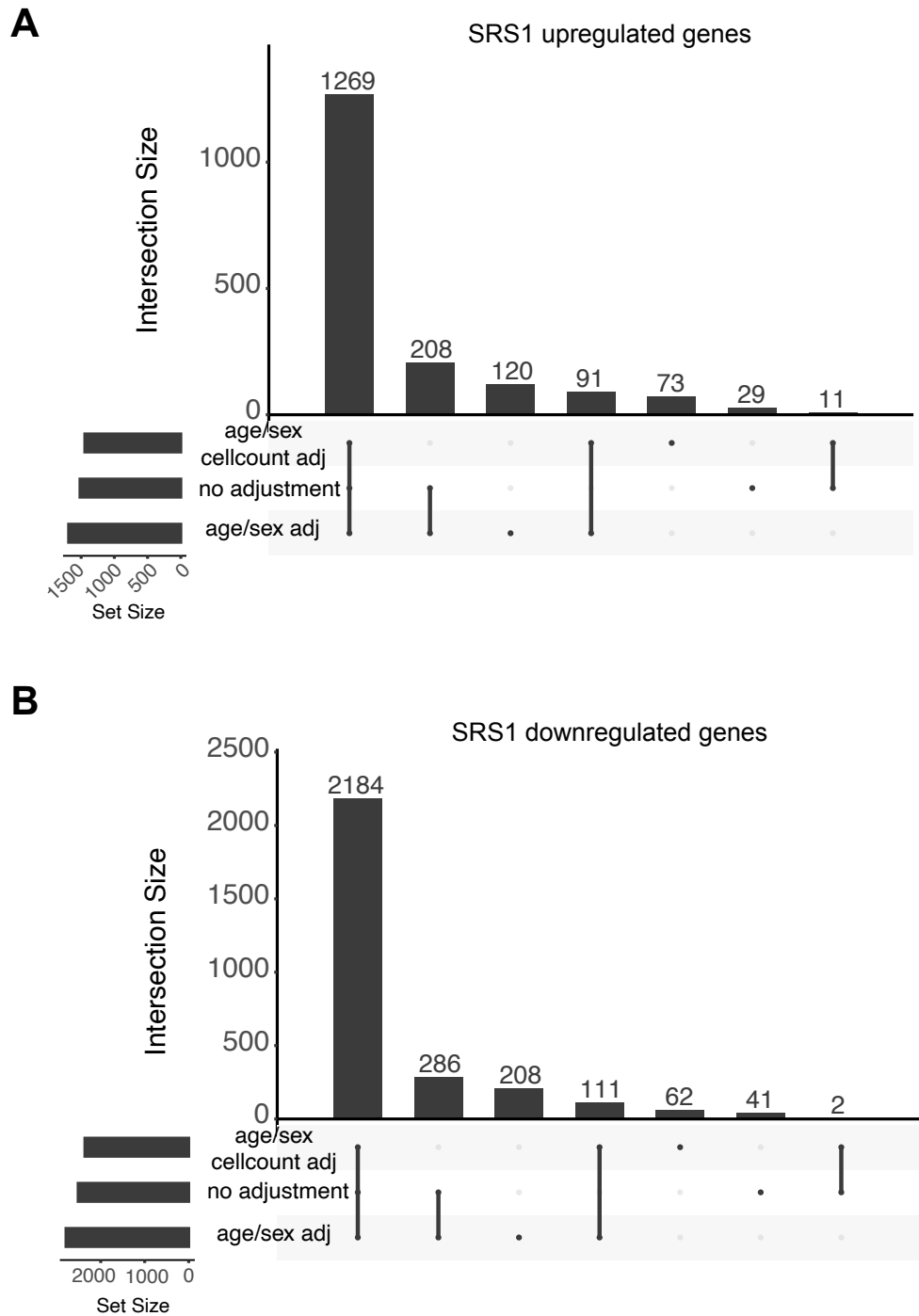
**Figure 3.14: Pathway analysis for sepsis vs. healthy volunteer (HV) differentially expressed genes.** (A-B) Gene Ontology (GO) biological processes (BP) pathways significantly enriched (false discovery rate <0.05) in (A) genes upregulated in sepsis-HV contrast and (B) genes downregulated in sepsis-HV contrast with 95% confidence intervals for odds ratios depicted. nOverlap = number of differentially expressed genes present in gene set.

hematology cell count data and included neutrophil and monocyte counts into the linear model on top of age and sex, with 3803 genes continuing to reach significance (Fig. 3.15B). Overlapping the three gene sets, I found 3453 genes to be either differentially up or downregulated regardless of the model specified, suggesting again that the majority of the SRS signal was deriving from gene expression shifts rather than from either baseline covariates such as age or sex, or due to peripheral blood immune cell abundance changes (Fig. 3.16A-B).

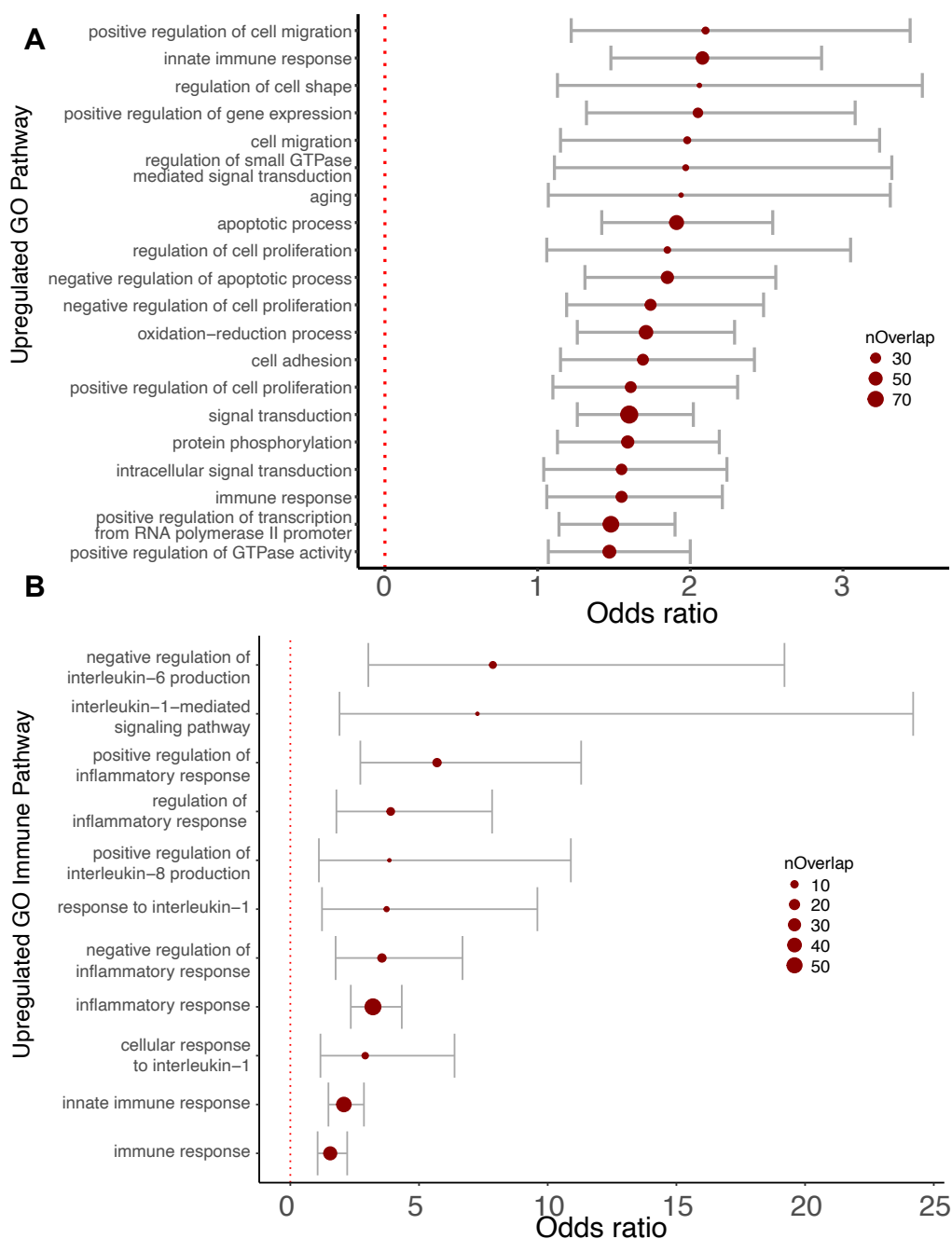


**Figure 3.15: Differential gene expression (DGE) of sepsis samples across sepsis response signatures (SRS).** (A-B) Volcano plots for DGE of sepsis samples across SRS (n=26 SRS1, n=16 non-SRS1), (A) accounting for age and sex covariates and (B) additionally including clinical peripheral blood neutrophil/monocyte cell counts into the DGE model. adj = adjusted, FC = fold change, Differentially expressed = FC  $\geq 1.5$  and adjusted p-value  $< 0.05$

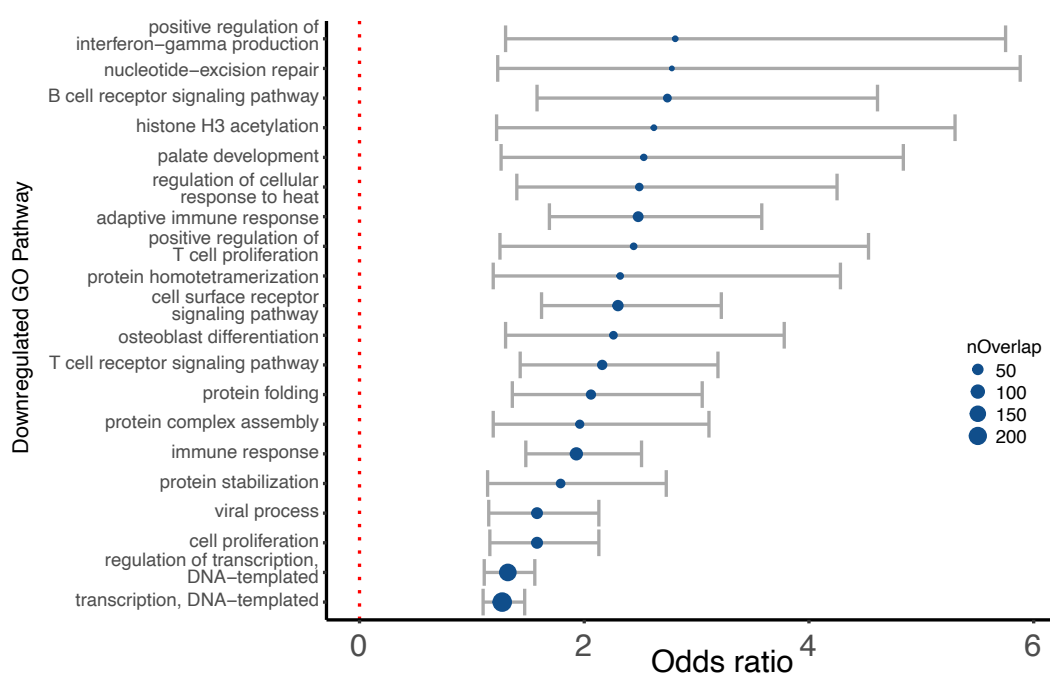
This comparison once again highlighted the general immune response pathway, this time enriched in genes upregulated in SRS1 (Fig. 3.17A). Moreover, there was a specific enrichment of the innate immune response. To further understand the immune pathways involved, I filtered for immune related pathways, finding a number of cytokine pathways to be enriched including response to IL-1, positive regulation of IL-8 and negative regulation of IL-6 (Fig. 3.17B). In addition, both B



**Figure 3.16: Differential gene expression (DGE) overlaps across sepsis response signatures (SRS) with and without covariates.** (A-B) Overlaps in (A) upregulated and (B) downregulated genes when comparing SRS1 vs. non-SRS1 sepsis patients between including vs. not including age, sex and clinical peripheral blood white cell count covariates.



**Figure 3.17: Pathway analysis for sepsis response signature (SRS) 1 upregulated genes in sepsis samples.** (A-B) Gene Ontology (GO) biological processes (BP) pathways significantly enriched (false discovery rate <0.05) in genes upregulated in SRS1 with 95% confidence intervals for odds ratios depicted. (A) depicts top pathways by odds ratio while (B) depicts pathways filtered for immune processes. nOverlap = number of differentially expressed genes present in gene set.



**Figure 3.18: Pathway analysis for sepsis response signature (SRS) 1 downregulated genes in sepsis samples.** Gene Ontology (GO) biological processes (BP) pathways significantly enriched (false discovery rate <0.05) in (A,B) genes downregulated in SRS1 with 95% confidence intervals for odds ratios depicted. nOverlap = number of differentially expressed genes present in gene set.

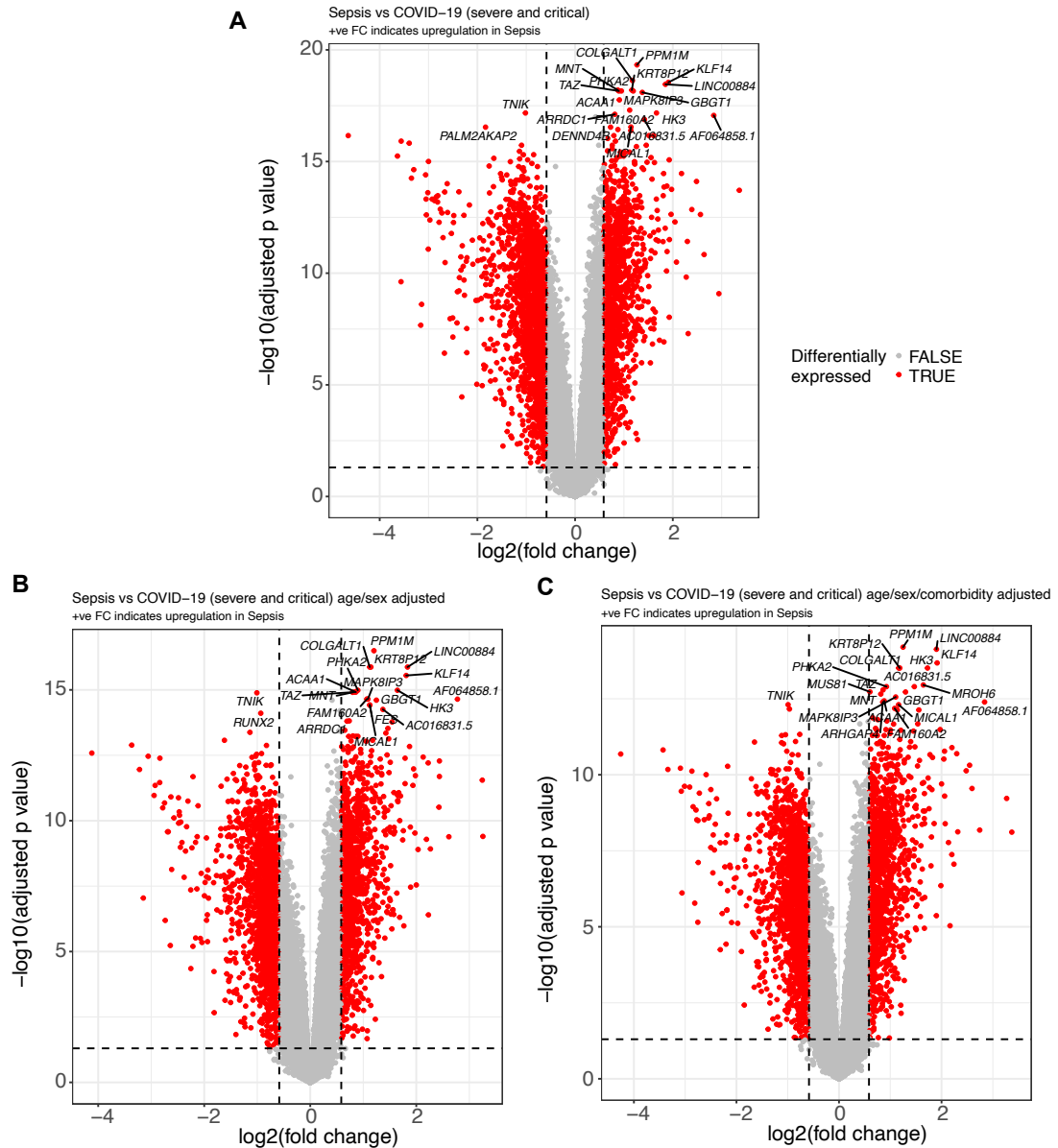


and T cell receptor signaling pathways were enriched in genes downregulated in SRS1, suggesting a component of reduced adaptive immunity in SRS1 (Fig. 3.18).

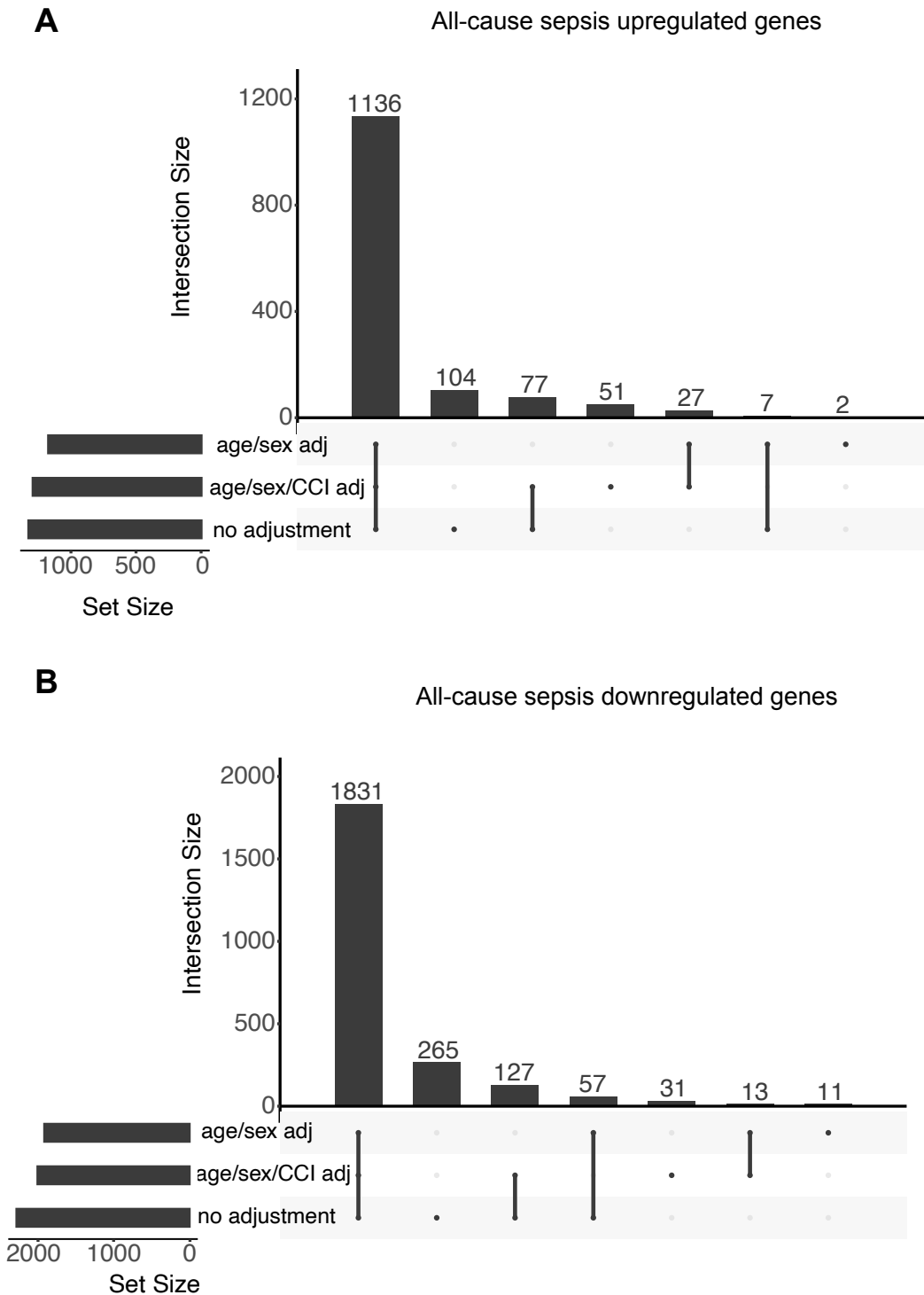
### **Differences between all-cause vs. COVID-19 viral sepsis**

I then compared between all-cause sepsis (n=42 samples which spanned severe and critical illness) and severe and critical COVID-19 patients (CS + CC, n=60 samples) to interrogate whether there were features specific to the host response to COVID-19. The list of DEGs stood at 3604 genes before age/sex adjustment, and 3084 genes after adjustment (Fig. 3.19A-B). In view of the possibility of differences in pre-morbid status of the patients accounting for some of the differences, I also included the Charlson co-morbidity index (CCI) in a further comparison, finding 3293 genes to be DE (Fig. 3.19C). Of the 1324 genes upregulated in all-cause sepsis without covariates in the linear model, 1213 i.e. 92% were also upregulated after accounting for age, sex and the CCI (Fig. 3.20A). Similarly, 1958/2280 i.e. 86% of downregulated genes were shared between the two models (Fig. 3.20B).

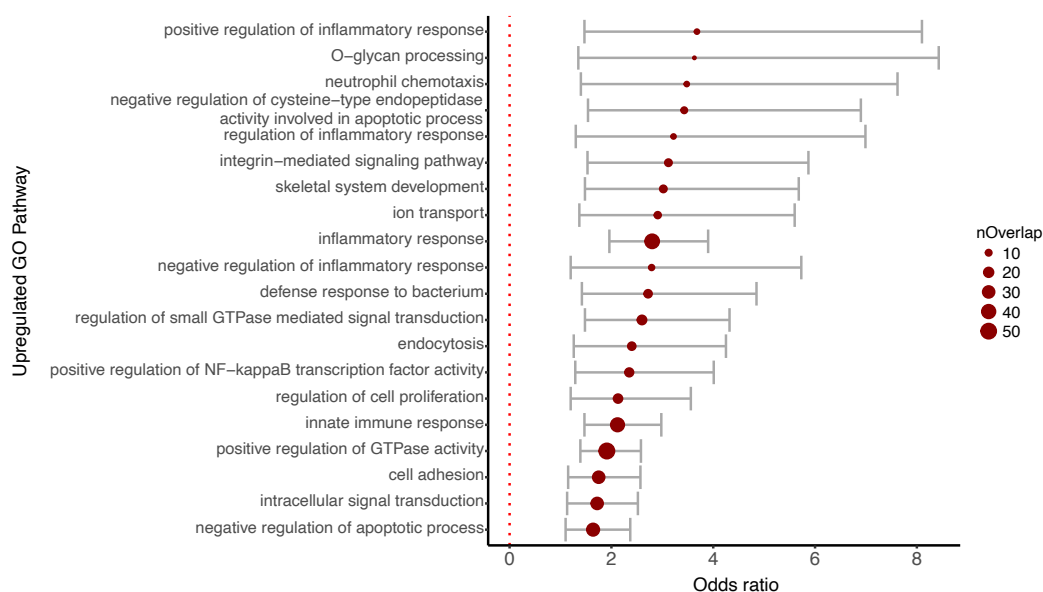
Pathway analysis results showed that there was an enrichment of neutrophil chemotaxis, as well as regulation (both positive and negative) of the inflammatory response in genes upregulated in all-cause vs. COVID-19 sepsis (Fig. 3.21A). In the downregulated genes, there was enrichment of pathways for blood coagulation (discussed below) and the viral process, likely due to the mostly bacterial causes in all-cause sepsis vs. COVID-19 which is caused specifically by the virus SARS-CoV-2 (Fig. 3.22A). When filtering for immune pathways, the pathway with the highest odds ratio was immunoglobulin production, and the pathway for the humoral response was also enriched, suggesting a stronger B cell response in COVID-19 viral sepsis (Fig. 3.22B). There was also enrichment of pathways related to IL-4, again possibly a reflection of the B cell response in COVID-19 viral sepsis. Both type I interferon



**Figure 3.19: Differential gene expression (DGE) of all-cause vs. COVID-19 sepsis** (A-C) Volcano plots for DGE of all-cause (n=42) vs. COVID-19 (n=60) sepsis, (A) not accounting for any covariates, (B) accounting for age and sex covariates and (C) additionally including the Charlson comorbidity index (CCI) into the DGE model. FC = fold change, Differentially expressed = FC  $\geq 1.5$  and adjusted p-value  $< 0.05$



**Figure 3.20: Differential gene expression (DGE) overlaps of all-cause vs. COVID-19 sepsis with and without covariates.** (A-B) Overlaps in (A) upregulated and (B) downregulated genes when comparing all-cause vs. COVID-19 sepsis between including for vs. not including age, sex and Charlson comorbidity index (CCI) covariates.

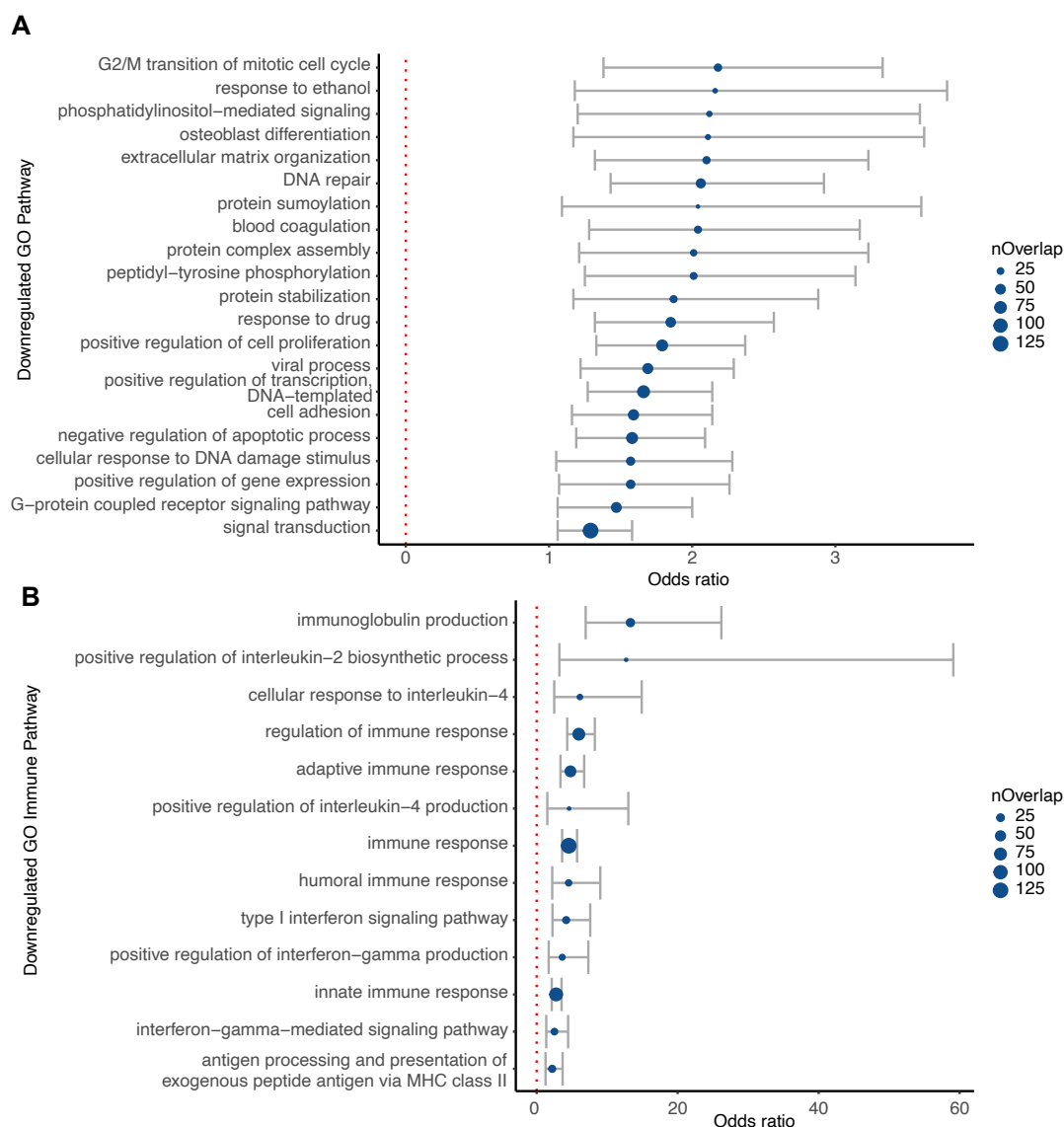


**Figure 3.21: Pathway analysis for genes upregulated in all-cause vs. COVID-19 sepsis.** Gene Ontology (GO) biological processes (BP) pathways significantly enriched (false discovery rate  $<0.05$ ) in genes upregulated in all-cause vs. COVID-19 sepsis. with 95% confidence intervals for odds ratios depicted. nOverlap = number of differentially expressed genes present in gene set.

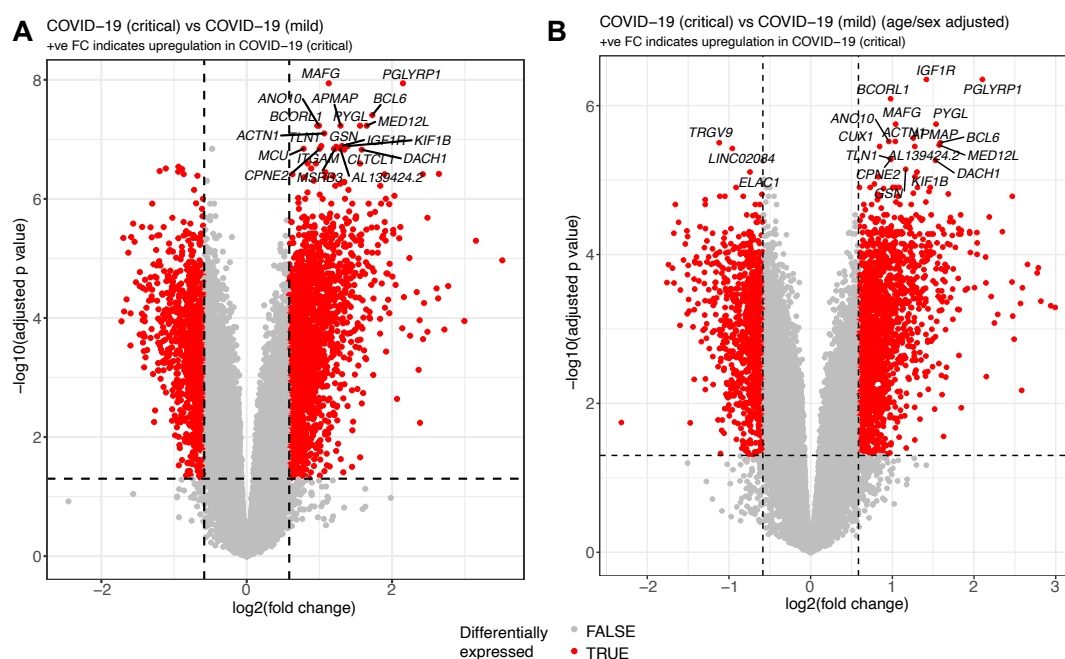
signaling pathways and interferon gamma related pathways were also enriched. The enriched immune pathways gave insight into the non-immune pathways enriched e.g. positive regulation of transcription and gene expression (Fig. 3.22A), as active B cells/plasmablasts producing a lot of immunoglobulin would be transcriptionally highly active. Finally, a pathway related to antigen presentation via MHC class II was enriched in the genes downregulated in all-cause sepsis, suggesting that there may be even greater immunosuppression in all-cause vs. COVID-19 viral sepsis (Fig. 3.22B).

### COVID-19 viral sepsis vs. non-septic COVID-19

For an alternative perspective of transcriptomic features of COVID-19 viral sepsis, I compared the sickest COVID-19 patients i.e. critically ill patients with those with mild disease. I found 3009 DEGs, again with the majority of these genes differentially expressed whether age, and sex covariates were included into the model (Fig. 3.23A-B, Fig. 3.24A-B). Pathways relating to angiogenesis and



**Figure 3.22: Pathway analysis for genes downregulated in all-cause vs. COVID-19 sepsis.** (A-B) Gene Ontology (GO) biological processes (BP) pathways significantly enriched (false discovery rate <0.05) in genes downregulated in all-cause vs. COVID-19 sepsis, with 95% confidence intervals for odds ratios depicted. (B) depicts pathways filtered for immune processes. nOverlap = number of differentially expressed genes present in gene set.

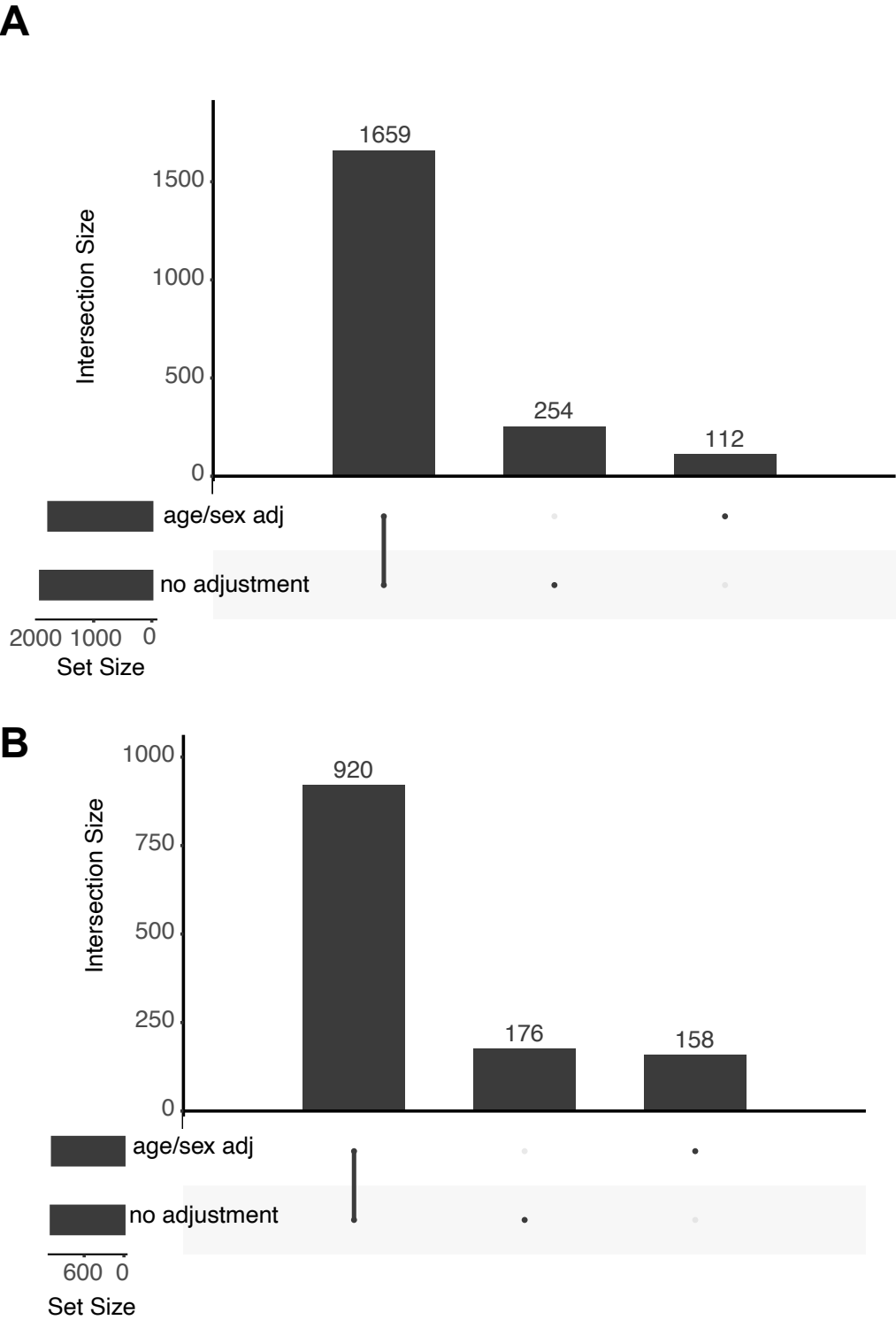


**Figure 3.23: Differential gene expression (DGE) of severe/critical vs. mild COVID-19 samples.** (A-B) Volcano plots for DGE of severe/critical vs. mild COVID-19 samples (A) not accounting for any covariates and (B) including age and sex covariates into the DGE model. FC = fold change, Differentially expressed = FC  $\geq 1.5$  and adjusted p-value  $< 0.05$

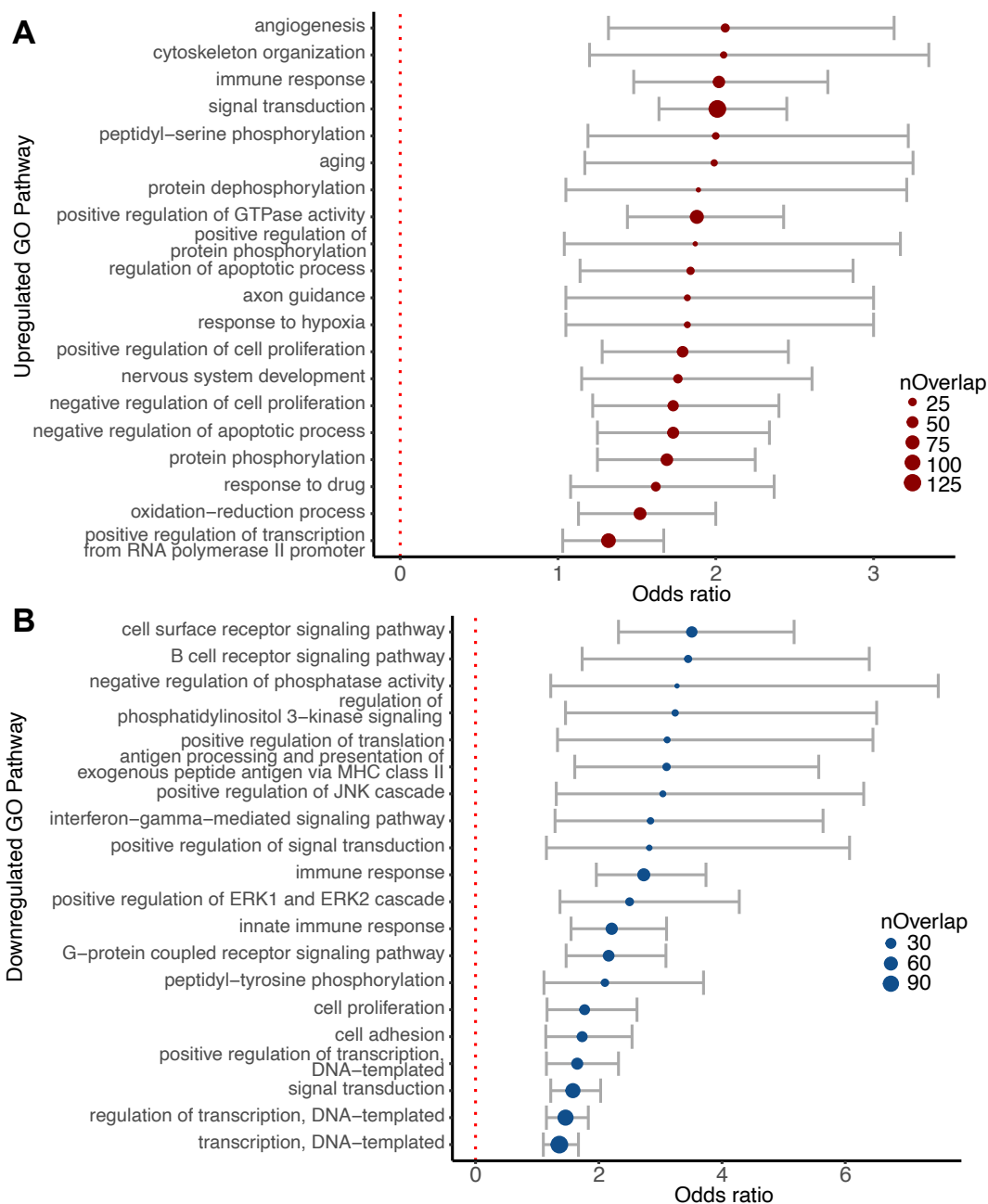
hypoxia were enriched in genes upregulated in critical COVID-19 (Fig. 3.25A). For genes downregulated in critical COVID-19, the B cell receptor signaling pathway was enriched, suggesting that there may be a lower B cell response in COVID-19 sepsis than the non-septic COVID-19 state (Fig. 3.25B) (discussed further below). There was also enrichment of the pathway for antigen processing and presentation via MHC class II in genes downregulated in critical COVID-19, highlighting the commonality of reduced MHC class II pathways in sepsis in general (Fig. 3.25B).

### 3.3 Discussion

The current definition of sepsis under Sepsis-3 implicates host response dysregulation to infection, but does not necessarily pinpoint specific mechanisms that are dysfunctional. The analyses I described in this chapter examined the bulk peripheral blood immune contributions which highlighted potential



**Figure 3.24: Differential gene expression (DGE) overlaps between critical vs. mild COVID-19 with and without covariates.** (A-B) Overlaps in (A) upregulated and (B) downregulated genes when comparing critical COVID-19 vs. mild COVID-19 including vs. not including age and sex covariates.



**Figure 3.25: Pathway analysis for critical vs. mild COVID-19 differentially expressed genes.** (A-B) Gene Ontology (GO) biological processes (BP) pathways significantly enriched (false discovery rate  $<0.05$ ) in (A) genes upregulated in critical COVID-19 and (B) genes downregulated in critical COVID-19 compared to mild COVID-19, with 95% confidence intervals for odds ratios depicted. nOverlap = number of differentially expressed genes present in gene set.



immune mechanisms and cellular contributions to all-cause and COVID-19 sepsis pathogenesis. In particular, I focused on the application of previously derived SRS groupings for stratifying sepsis patients to new cohorts of patients, and compared the similarities and differences in gene expression across all-cause sepsis and COVID-19 patient groups.

### **3.3.1 SRS as a unifying framework for sepsis**

One outstanding question of transcriptomic subphenotyping is the external validity of existing frameworks and definitions i.e. whether they can be applied in broader contexts than originally defined. Here, I show three aspects through which SRS can be extended; first, that SRS can be extended to non-ICU sepsis patients, second, that SRS applies for all-cause sepsis i.e. beyond the previously interrogated causes of CAP and FP, and third, that SRS can also be used in COVID-19 viral sepsis.

The applicability of SRS to COVID-19 is interesting given that while SARS-CoV-2 can cause viral sepsis, the majority of COVID-19 patients do not meet the clinical criteria for the definition of sepsis. Indeed, this is reflected by the mortality rates of the two conditions, where mortality rates of COVID-19 pre-vaccinations have been reported to be between 0.5-3% (Meyerowitz-Katz and Merone 2020; Mahase 2020; Verity et al. 2020), whereas the commonly observed mortality of sepsis is around 20-30%, and rising to close to 40% in the case of septic shock (Rhee et al. 2019; Bauer et al. 2020). Indeed, given the association of SRS with mortality, this is concordant with the low number of COVID-19 patients being assigned the most deleterious SRS1 state.

While the mortality from COVID-19 is low compared to sepsis, it is nonetheless high for a respiratory viral illness, and absolute death numbers are substantial due to the highly infectious nature of the virus. Compared to influenza which is known to cause pandemics and severely disrupt societies socioeconomically,

COVID-19 (pre-vaccination) is estimated to have a mortality rate of one to two orders of magnitude greater (Iacobucci 2021; Cozza et al. 2021). In particular, in the elderly age groups, COVID-19 mortality rates climb exponentially, going from 1% at 60 years to over 20% for 90 year olds (COVID-19 Forecasting Team 2022). This reflects the potential utility of applying SRS by the bedside for elderly and particularly ill COVID-19 patients, e.g. those above the age of 60 and/or patients requiring ICU admission.

Another interesting observation was the grouping of non-COVID-19 sepsis patients by SRS status rather than ICU status on hierarchical clustering analysis of the peripheral blood transcriptome. ICU admission is usually a clinical decision based on a combination of the presence of multi-organ failure, reversibility of the condition and absence of reasons for triaging the patient out of intensive care. On one hand, the separation of patients by SRS rather than ICU status may be reflecting the capture by SRS of a biological axis not easily detectable clinically; however, certain patients may have been triaged out of ICU admission with a general hospital ward ceiling of care. As such, while the results from the present analysis recapitulate the association of SRS and mortality, they do not necessarily reflect any gaps in adjudicating severity of illness in non-ICU SRS1 patients. Future studies including clinical data on the decided ceiling of care for each patient would be needed to understand the value of transferring non-ICU patients discovered to be SRS1 to the ICU.

### **3.3.2 Relevant disease mechanisms in COVID-19/sepsis**

The bulk RNA-seq data also provides a window into potential mechanisms relevant to COVID-19 and sepsis pathogenesis. Studies of COVID-19 peripheral blood immune mechanisms have mostly focused on comparing between varying COVID-19 severities, or occasionally with other conditions such as influenza (Lee et al. 2020). In addition, many of the reports, especially those where

single cell transcriptomics have been performed, focused on peripheral blood mononuclear cells (PBMCs) (Wilk et al. 2020; Zhu et al. 2020; Yao et al. 2021; Stephenson et al. 2021; Ren et al. 2021).

These studies have consistently demonstrated increased B cell / plasmablast responses with increasing severity of COVID-19, which is inconsistent with the bulk RNA-seq data presented above where B cell receptor signaling pathways were enriched in genes downregulated in critical vs. mild COVID-19. This contradiction could be accounted for by the fact that PBMC studies would miss global peripheral blood changes given the exclusion of granulocytes i.e. the proportion of B lymphocytes may be reduced in peripheral blood overall given the greater neutrophilia with more severe disease (Table 4.1). On the other hand, the increase in B cell / plasmablast responses with more severe COVID-19 on a per cell basis or within the mononuclear cell compartment that could be captured with the PBMC single-cell studies would not have been detected in the bulk transcriptomic data presented above as it does not have the single-cell resolution.

While the humoral response is elevated in COVID-19 viral sepsis vs. non-COVID-19 sepsis, Sweeney and colleagues (2021) did not find enrichment of similar pathways comparing COVID-19 against other viral illnesses, suggesting that elevated humoral responses may be reflective of an antiviral response generally rather than being specific to COVID-19 viral sepsis.

Here, comparisons between septic COVID-19 patients and non-COVID-19 sepsis samples also demonstrated that general inflammatory responses and neutrophil activity were elevated in all-cause vs. COVID-19 viral sepsis. Since the time of data generation and analysis for this chapter, Aschenbrenner et al (2021) have also performed a similar whole blood transcriptomic experiment and analysis as in this chapter. In that study, comparisons between severe COVID-19 and bacterial sepsis also showed enrichment in bacterial sepsis of a neutrophil module, consistent with the current findings.

Another notable mechanism highlighted in the analysis above was the enrichment of the coagulation pathway in COVID-19 viral sepsis vs. all-cause sepsis, consistent with the thrombotic tendencies that are well described in COVID-19 (Malas et al. 2020; Loo et al. 2021). This may also be related to the enrichment of angiogenesis in critical over mild COVID-19 as angiogenic factors e.g. VEGF can be pro-thrombotic (Clauss et al. 1996; Zanetta et al. 2000). However, due to the lack of cell type resolution in bulk RNA-seq data, it is not possible from this dataset to understand where the signal for angiogenesis and coagulation may be coming from. An association of COVID-19 severity with megakaryocytes has been shown (Stephenson et al. 2021), implicating them as a potentially relevant cell type.

In terms of mechanisms underlying sepsis heterogeneity, the DE testing and pathway enrichment analyses across SRS groups showed features of heightened inflammation in SRS1 in the form of increased cytokine signaling pathways, but at the same time reinforced the notion of immunosuppression in severe sepsis (Poll et al. 2021) as seen by the downregulated B and T cell pathways in SRS1. This may be related to the increased apoptosis of lymphocytes in sepsis (Hotchkiss et al. 1999; Chang et al. 2007; Shankar-Hari et al. 2017; Boomer et al. 2011). T lymphocyte exhaustion has been reported in sepsis previously (Boomer et al. 2011) and in SRS1 (Davenport et al. 2016), but pathway analysis from the current dataset did not directly point to this phenomenon. Notably, T cell exhaustion in SRS1 has not been experimentally validated. Whether the previously described T cell exhaustion in SRS1 matches with the agreed upon understanding of T cell exhaustion in terms of a hierarchical loss of T cell function with increased inhibitory receptor expression, reduced cytokine secretion and proliferative potential as well as characteristic enduring epigenetic changes (Blank et al. 2019) remains to be more formally tested. This could be clarified by in vitro T cell interrogation of cytokine secretion and proliferation

and epigenetic assays to look for characteristic changes e.g. around the *PDCD1* locus (Pauken et al. 2016; Sen et al. 2016) in SRS1 T cells.

### **COVID-19 and the interferon response**

Meanwhile, the relationship between antiviral type I interferons and COVID-19 severity remains contentious with early reports using in vitro models of SARS-CoV-2 infection showing that the coronavirus shut interferon signaling down relative to other viruses (Blanco-Melo et al. 2020). However, analysis of patient respiratory tract samples showed robust interferon stimulated gene expression (Zhou et al. 2020). Iwasaki and colleagues (2020) have reported a correlation between nasopharyngeal viral load with plasma interferon levels, suggesting that the heightened levels of interferon are not successful in controlling viral replication and may therefore even have a pathogenic role as has previously been described with SARS-CoV (Cameron et al. 2007). A possible explanation for the varying observations is that the interferon response is spatially highly specific, as it has been shown there is differential interferon response in COVID between upper and lower airways with type III interferon elevation in upper airways specifically associated with good outcomes (Sposito et al. 2021). In the current bulk RNA-seq dataset, it is informative that while the type I interferon pathway is elevated in COVID-19 vs. non-COVID-19 sepsis, likely reflecting the viral nature of the illness, I did not find a similar enrichment in this pathway comparing critical against mild COVID-19 samples. This is reinforced by the findings from Aschenbrenner et al (2021) where interferon responsive genes were observed irrespective of COVID-19 severity.

### **3.3.3 Limitations**

While the current dataset enables insights to be drawn on shared and distinct aspects of COVID-19 and non-COVID-19 sepsis, there is nonetheless a lack of comparator groups of other viral illnesses, especially those which can lead to

sepsis e.g. influenza. As most hospital cases of sepsis are due to bacterial pathogens (Mayr et al. 2014), this means that the analysis to a certain extent reflects a bacterial vs. viral comparison. The present analysis does not delineate the specific aspects COVID-19 viral sepsis that are distinct from other forms of viral sepsis.

Additionally, given the natural history of COVID-19 caused by the original strain of SARS-CoV-2 where deterioration often occurs 7-14 days after symptom onset (Huang et al. 2020; Shah et al. 2020, Burn et al. 2021), it was inevitable that the septic COVID-19 samples in the dataset reported were further away from symptom onset than the all cause sepsis samples, to a certain extent confounding the analysis. More detailed matching of clinical features across comparator groups would enhance the quality of the dataset, although especially in the setting of a pandemic and overwhelmed healthcare system, would be logistically profoundly challenging.

Another limitation of the current approach is that bulk transcriptomics represents the average gene expression of the entire sampled biological tissue material and therefore loses cell type resolution. Not only does this obscure how particular cell types may be behaving across the different conditions, it is also not possible to differentiate whether changes seen are due to changes in cellular composition, changes in gene expression within the same cell types, or as is biologically likely to be the case, a combination of the two. Cell type deconvolution algorithms exist to deconvolute bulk transcriptomic data to give cellular abundance estimates (Newman et al. 2015; Wang et al. 2019, Zaitsev et al. 2019) and in some cases even the gene expression profiles of each cell type or state that is deconvoluted (Newman et al. 2019; Dong et al. 2021; Chu et al. 2022). However, the vast majority of these methods require some reference ground truth dataset either in the form of pure bulk RNA-seq profiles or often single cell RNA-seq data, or even in the rare instance of not requiring a ground

truth e.g. by Zaitsev et al (2019), has been shown to be significantly inferior in performance to methods utilising a reference (Sutton et al. 2022). The lack of whole blood single cell RNA-seq datasets preserving granulocytes and in the exact disease context of COVID-19 and sepsis therefore precluded such a method to circumvent the lack of cellular resolution in the current data.

Single cell analyses are therefore required for this which are described in the subsequent chapters. One mitigating effect in bulk transcriptomics is to take into differential cell counts where possible, which was performed when contrasting across SRS groups, showing that the changes in gene expression were likely not only due to cell composition changes.

### **3.4 Conclusions**

The analysis presented in this chapter examined transcriptomic data from a number of infection groups, ranging from all-cause sepsis patients, sepsis from COVID-19, as well as various control groups including healthy volunteers and patients with COVID-19 but who did not suffer from sepsis. The analysis has highlighted the shared and distinct aspects of the COVID-19 vs. all-cause sepsis transcriptomic responses. Given the applicability of SRS in terms of the host transcriptomic response to both viral COVID-19 sepsis and non-COVID-19 all-cause sepsis, similar treatment strategies could be considered for both groups of patients based on their underlying SRS profile.

# A LARGE SCALE SINGLE-CELL MULTI-OMIC PBMC ATLAS OF SEPSIS, SEVERE COVID-19 AND INFLUENZA

*This chapter explores the peripheral blood mononuclear cell changes in COVID-19, severe influenza and sepsis, comparing across conditions to identify shared and distinct aspects of the immune response.*

|     |                        |     |
|-----|------------------------|-----|
| 4.1 | Introduction . . . . . | 108 |
| 4.2 | Results . . . . .      | 116 |
| 4.3 | Discussion . . . . .   | 146 |
| 4.4 | Conclusions . . . . .  | 151 |

## 4.1 Introduction

### 4.1.1 Clinical cohorts

In this chapter, I present data generated and analyzed together with other members of the COMBAT consortium from samples of the same cohort as the previous chapter (healthy volunteers (HVs), patients hospitalized with mild, severe and critical COVID-19 infection (CM, CS and CC), patients in the community with mild COVID-19 infection (CComm), as well as all-cause sepsis patients (Sepsis)). In addition, I included samples from individuals with severe influenza (flu) requiring intensive care unit (ICU) admission (COMBAT



consortium 2022). When the severity of either COVID-19 or influenza is sufficiently high, patients meet the definition of organ dysfunction secondary to a dysregulated host response to infection according to the Sepsis-3 criteria.

This permitted a comparison between the conditions of all-cause sepsis, COVID-19 viral specific sepsis and flu viral sepsis to move beyond the understanding derived from bulk RNA-seq in the previous chapter. Comparing between COVID-19 and influenza, specificities of the host response to COVID-19 can be further isolated beyond identifying only antiviral components when comparing to all-cause sepsis.

As per chapter 3, the data in this chapter have also been published as part of the COMBAT consortium work (COMBAT consortium 2022).

#### **4.1.2 Cell theory and single cell profiling technologies**

Formulated in the mid-19th century, cell theory laid the conceptual foundations for modern biology. The theory states that 1) all living organisms are composed of one or more cells 2) the cell is the most basic unit of life and 3) all cells arise from pre-existing cells. As the cell is the basic unit of life, understanding the biology of individual cells is of paramount importance for understanding the biology of an organism (Harris 2000).

Since flow cytometry was first developed in the late 1960s, single cell analysis has been an indispensable part of the toolkit of any immunological investigation (McKinnon 2018). Single cell profiling technologies have seen rapid advancements in recent years which allow individual cells to be interrogated with substantially higher parameterisation than flow cytometry previously allowed. While improvements in fluorescence and laser technologies have allowed flow cytometry to commonly interrogate over 10 markers per experiment, limitations nevertheless exist due to fluorophore emission spectra overlap. Mass cytometry (cytometry by time of flight, CyTOF) instead uses antibodies conjugated to

rare earth metals rather than fluorophores and detection by mass spectrometry (Bendall et al. 2011; Spitzer and Nolan 2016). This has allowed antibody panels to easily scale to upwards of 40 markers. The key drawbacks are the loss of the cell during the experiment (precluding cell sorting) and a significantly slower acquisition time than flow cytometry.

While the cytometric techniques focus on protein detection on the cell surface or intracellularly, gene expression has also long been used as a readout for biological investigation, as in chapter 3. Obtaining such a readout from individual cells was pioneered as far back as 1990 (Brady et al. 1990; Van Gelder et al. 1990) with selected genes probed by Southern hybridization, which eventually progressed to readouts of hundreds (Chiang and Melton 2003) and then thousands of genes (Kamme et al. 2003) with microarray hybridization. Commercialized platforms, as applied to neural (Tietjen et al. 2003) and bone marrow tissues (Seshi et al. 2003), followed soon after, and the transition beyond hybridization towards sequencing was made via an intermediate step of Sanger sequencing in 2006 (Moroz et al. 2006) before next generation sequencing technologies were adapted for single cell transcriptomic use in 2009 (Tang et al. 2009).

The past decade has then seen a striking rise in the prominence, accessibility and scale of single cell transcriptomics (scRNA-seq), enabled chiefly by microfluidic technologies (Macosko et al. 2015; Jovic et al. 2022). Compared to cytometry, scRNA-seq is significantly more expensive per cell and experiments therefore interrogate far fewer cells, but the increase from a single digit number of cells when the technology was first invented to upwards of tens of thousands of cells has meant that the technology is no longer an esoteric method of purely academic interest. This has allowed the complex cellular milieu to be deconvoluted in an unprecedented fashion in both scale and resolution.

Further advances in the field of single cell genomics have happened in the recent past, with techniques enabling the interrogation of cell surface proteins

in single cell sequencing experiments. Developed in 2017, cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) combines scRNA-seq with a sample staining step with oligo-attached antibodies used to label cell surface proteins (Stoeckius et al. 2017). The oligos are then captured along with the RNA in downstream steps, sequenced, and deconvoluted to produce a readout for cell surface protein expression levels of markers of interest. This has been of particular interest in the field of systems immunology, as it has been shown that identifying known immune cell types purely based on transcriptomics has been especially difficult within the T cell compartment (Zheng et al. 2017; Hao et al. 2021). A wide array of other technologies now exist to profile multiple modalities in individual cells. This includes SHARE-seq (Ma et al. 2020) and scNMT-seq (Clark et al. 2018) which respectively profile RNA and chromatin accessibility simultaneously, versus RNA, methylation and histone marks simultaneously.

### 4.1.3 Cell atlasing

The advent of high-throughput single cell transcriptomics has had far reaching implications for biology, with efforts towards deep molecular characterization of cell types and states leveraging this technology in an attempt to produce "cell atlases". The Human Cell Atlas (HCA) has been a major international collaborative scientific initiative to comprehensively catalogue all cell types in the estimated 37 trillion cells in the human body, and ultimately to provide a cellular reference map, much like the rationale for the Human Genome Project in sequencing the whole human genome to provide a reference for future studies of the mechanisms underlying human genetic variation in health and disease (Regev et al. 2017; Lindeboom et al. 2021). Indeed, the importance of the cell as a relevant level of biological abstraction has long been championed by Sydney Brenner (Brenner 2010), and his original idea of Cellmap has now seen life in the form of the HCA. An example of progress in these efforts is the extensive catalogues of cell types and states detailed through projects such as the

Tabula Sapiens initiative, where over 500,000 cells spanning 400 cell types from humans were profiled to provide cell specificities in splicing profiles and cell type similarities across tissues amongst other findings (Tabula Sapiens Consortium\* et al. 2022).

Ultimately, the single cell resolution is such that the transcriptomic readout is within a preserved endogenous context which is lost in bulk transcriptomics. In the context of immunity, the current possible scale of scRNA-seq allows a near-saturating analysis of all immune cell types, allowing an exhaustive and quantitative description of cellular diversity beyond what was previously possible. While immune cells have previously often been defined by selected surface markers, scRNA-seq represents a data-driven and unbiased alternative where commensurate computational advances to analyze the large amounts of scRNA-seq data have allowed us to systematically characterize and classify cell types and states.

In this chapter, I present a large scale atlas dataset of joint cell surface protein (antibody derived tags, ADT) and whole transcriptome profiling on individual peripheral blood mononuclear cells (PBMCs) across the various conditions of severe infection.

#### **4.1.4 PBMCs as a relevant tissue type**

Peripheral blood mononuclear cells (PBMCs) are immune cells with a single, round nucleus, which includes monocytes, lymphocytes, dendritic cells, NK cells and hematopoietic stem and progenitor cells. In the laboratory, these cells are isolated from density gradient centrifugation of whole blood (e.g. with Ficoll, a medium with a density of 1.077 g/mL) where granulocytes and red cells are found in the high density fraction and can be removed for PBMC purification.

PBMCs are a relevant sample for the study of sepsis given the immune nature of the dysregulation of the host response that characterises sepsis. In addition,

PBMCs are robust to cryopreservation, permitting biobanking followed by batch profiling which is key in studies of patient cohorts in order to mitigate batch effects and to allow careful experimental design to avoid confounding.

The key drawback to profiling PBMCs is that granulocytes, which comprise more than half of the leukocytes in peripheral blood, are not interrogated. However, important logistical considerations make granulocyte profiling difficult, namely the inability to store granulocytes, necessitating work with fresh samples which while possible on smaller scales ultimately limits sample size.

Although there are limitations, the use of PBMCs to study severe infections should not be underestimated. The study of PBMCs in sepsis has provided us with the understanding of the association of lymphopenia and reduced monocyte HLA-DR expression with sepsis severity (Drewry et al. 2014; Monneret et al. 2006), and flow cytometric immunophenotyping of PBMCs in COVID-19 has shown differences in T cell responses between male and female patients (Takahashi et al. 2020). Since the time of data generation and analysis, a number of other studies have examined PBMCs with single-cell -omics similar to the dataset presented in this chapter (Wilk et al. 2020; Stephenson et al. 2021; Zhu et al. 2020; Ren et al. 2021), the findings of which are discussed below in the context of the results of this chapter. In the context of flu, studies of PBMCs have revealed differences in T cell populations across age groups e.g. an increase in CD45RA effector memory subsets in older age groups (Wagar et al. 2011), and examined T cell responses to flu vaccines (Powell et al. 2013). In total, this suggests that substantial value can be derived from PBMC profiling in sepsis and COVID-19/flu.

#### **4.1.5 Influenza**

One key extra patient comparator group not included in the previous chapter is a cohort of patients with severe influenza requiring ICU admission. Flu is

an infectious respiratory disease much like COVID-19, and is caused by either influenza A or B viruses in humans, which are negative-sense single stranded RNA viruses (Krammer et al. 2018). In addition to annual seasonal epidemics, sporadic flu pandemics occur, often due to zoonotic influenza A viruses, such as the infamous "Spanish flu" during and after the first World War, and more recently, an H1N1 swine flu pandemic in 2009. The "H" and "N" naming of flu strains derives from the variant of the hemagglutinin (HA) and neuraminidase (NA) viral proteins, which respectively mediate viral entry into and exit from host cells.

It is estimated that flu results in around 500,000 deaths a year (Lozano et al. 2012). At the cellular level, influenza virus mainly affects respiratory epithelial cells by HA binding with host cell sialic acid molecules, and as such usually causes a respiratory illness with nonspecific systemic symptoms such as fever, headaches, myalgia and malaise (Krammer et al. 2018). The host immune response to flu follows the stereotyped patterns of viral responses where the innate immune system detects influenza virus pathogen associated molecular patterns (PAMPs) through pattern recognition receptors (PRRs) such as retinoic acid-inducible gene I protein (RIG-I) detecting viral RNA (Pichlmair et al. 2006). This leads to a type I IFN response with induction of IFN inducible antiviral genes such as *MX1* and *OAS1* (Rajsbaum et al. 2012). The adaptive immune system is then recruited with clonal selection and expansion of T and B lymphocytes to ultimately clear the virus. While most patients clear the virus, some develop severe disease either due to the virus, or secondary bacterial infection of the lower respiratory tract, which can ultimately lead to sepsis.

#### **4.1.6 Modular gene expression analysis - WGCNA**

As discussed in chapter 1, one of the pitfalls of standard differential gene expression (DGE) analysis is the loss of statistical power from the heavy multiple

testing burden when checking for differential expression of each detected gene. In addition to this, the sparsity of scRNA-seq data makes it even more difficult for DGE analysis to achieve adequate power (Crowell et al. 2020), motivating the development or use of other tools.

One workaround is therefore to analyze gene expression in units of genes e.g. modules or programs. A method to construct such modules is through weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath 2008), where genes are correlated with one another and the co-expression then weighted by a thresholding value in order to obtain a network which reaches scale-free topology (Zhang and Horvath 2005). This is followed by unsupervised hierarchical clustering of genes into modules, which can then be summarized by eigengene expression i.e. the first principal component of the module. Eigengene expression can then be tested per module across different conditions, mitigating the multiple comparisons of testing each individual gene across every condition. This method will be deployed as the method of choice in DGE analysis for the CITE-seq dataset presented.

#### **4.1.7 Aims**

The overall aim of this chapter is to provide a detailed atlas level bioresource catalogue of all cell types and cell states in the PBMC compartment in health and in all-cause/COVID-19/influenza sepsis. I will look for shared and distinct aspects of the response to all-cause vs. COVID-19 specific sepsis. Specifically, I aim to:

1. generate a single-cell multimodal (CITE-seq) atlas of the immune response in PBMCs in sepsis traits and comparator groups, with opportunities for future reanalysis
2. establish the differential response at the single-cell resolution (both cell abundance and gene expression) between all cause and viral sepsis, and

between COVID-19 and influenza as causes of viral sepsis

3. understand the biological processes involved in the differential response between sepsis comparator groups by performing co-expression network analysis

## 4.2 Results

### 4.2.1 Experimental approach and design

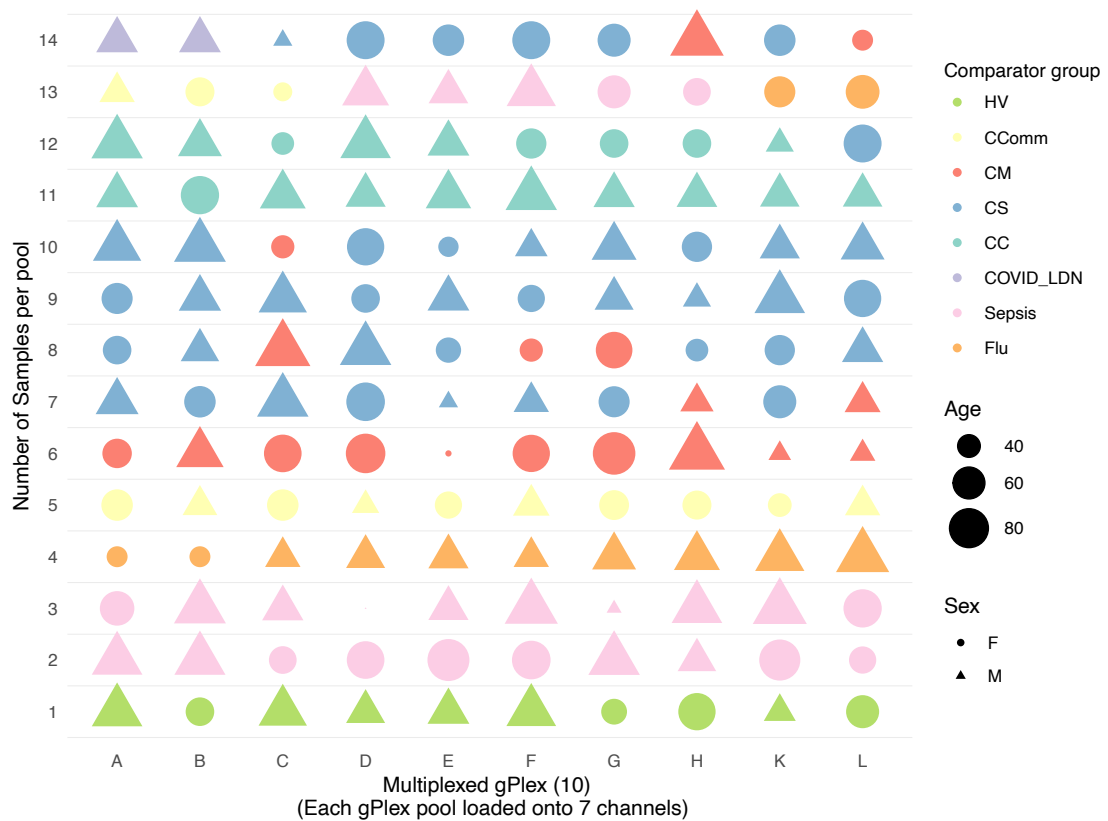
The patient cohorts in this chapter very closely align with chapter 3. The same set of 78 hospitalised and 13 community COVID-19 samples and 25 of the 44 sepsis samples were included. The healthy volunteer (HV) samples were also the same as in the previous chapter. In addition, I included a further 12 samples of acute severe influenza requiring ICU admission to St. George's Hospital, as well as two samples from patients critically ill with COVID-19 also admitted to ICU at St. George's Hospital.

In total, the dataset presented includes n=140 samples (n=124 unique individuals) from HVs, patients hospitalised with mild, severe, or critical COVID-19 (CM, CS, CC), patients in the community with mild COVID-19 (CComm) and non-COVID-19 all cause sepsis patients for single cell sequencing of PBMCs (Table 4.1). The samples were split into 10 pools, with each pool run across seven channels of a 10X Genomics chip for single cell capture. Cryopreserved PBMCs were thawed, pooled, FACS sorted for viability (with 7-AAD staining), followed by surface antibody staining with a 192 CITE-seq antibody panel. 50,000 cells were loaded per channel, aiming for a capture of 4000 cells per sample. Details of the workflow and pool design are shown in Fig. 4.1 and Fig. 4.2 .



| Trait                                                         | HV (n=10)    | CComm (n=13) | CM (n=17)      | CS (n=28)      | CC (n=19)      | Sepsis (n=23)   | Flu (n=12)   | COVID.London (n=2) | FDR     |
|---------------------------------------------------------------|--------------|--------------|----------------|----------------|----------------|-----------------|--------------|--------------------|---------|
| Age, median (range)                                           | 60.5 (42-76) | 49 (34-59)   | 68 (21-94)     | 57.5 (22-78)   | 57 (37-78)     | 70 (20-89)      | 56.5 (37-82) | 57 (56-58)         | 0.1     |
| Male sex, n (%)                                               | 6 (60%)      | 5 (38%)      | 8 (47%)        | 16 (57%)       | 14 (74%)       | 14 (61%)        | 8 (67%)      | 2 (100%)           | 0.46    |
| SOFA score, median (range)                                    | -            | -            | 0 (0-1)        | 2 (0-4)        | 7 (4-13)       | 3 (0-12)        | -            | -                  | <0.0001 |
| SaO <sub>2</sub> /FiO <sub>2</sub> , median (range)           | -            | -            | 4.6 (4.4-4.8)  | 2.6 (1.0-4.0)  | 1.9 (1.1-2.7)  | 2.4 (0.9-4.5)   | -            | -                  | <0.0001 |
| Charlson comorbidity index, median (range)                    | -            | -            | 0 (0-27)       | 4 (0-39)       | 0 (0-16)       | 0 (0-31)        | -            | -                  | 0.23    |
| Estimated days from symptom onset to sampling, median (range) | -            | 14 (7-22)    | 6 (3-26)       | 11.5 (2-25)    | 13 (3-25)      | 5 (0-15)        | 10.5 (1-25)  | 14 (13-15)         | 0.081   |
| Length ICU stay, median (range)                               | -            | -            | 0 (0-0)        | 8 (2-21)       | 23 (8-56)      | 7 (1-37)        | 22 (4-82)    | 13.5 (8-19)        | 0.00055 |
| Ventilation days, median (range)                              | -            | -            | 0 (0-0)        | 0.5 (0-33)     | 21 (7-49)      | 1 (0-12)        | 13.5 (2-63)  | 13 (8-18)          | <0.0001 |
| Ever vasopressors, n (%)                                      | -            | -            | 0 (0%)         | 4 (14%)        | 17 (89%)       | 8 (35%)         | 12 (100%)    | 2 (100%)           | <0.0001 |
| Ever renal replacement therapy, n (%)                         | -            | -            | 0 (0%)         | 2 (7%)         | 6 (32%)        | 4 (17%)         | 6 (50%)      | 1 (50%)            | 0.0071  |
| Death, n (%)                                                  | -            | 0 (0%)       | 0 (0%)         | 4 (14%)        | 3 (16%)        | 6 (26%)         | 2 (17%)      | 1 (50%)            | 0.026   |
| Laboratory findings, median (range)                           | -            | -            | 7.0 (2.2-11.2) | 8.3 (3.8-15.2) | 8.5 (3.7-17.4) | 12.9 (1.5-37.6) | -            | -                  | 0.0056  |
| Leukocytes, x10 <sup>9</sup> /L                               | -            | -            | 4.8 (1.1-8.3)  | 5.8 (1.7-13.4) | 7.5 (2.7-16.5) | 10.8 (1.2-34.1) | -            | -                  | 0.002   |
| Neutrophils, x10 <sup>9</sup> /L                              | -            | -            | 1.3 (0.6-2.5)  | 1.0 (0.3-2.3)  | 0.7 (0.3-1.8)  | 0.9 (0.2-1.9)   | -            | -                  | 0.026   |
| Lymphocytes, x10 <sup>9</sup> /L                              | -            | -            | 0.5 (0.1-1.4)  | 0.5 (0.2-1.1)  | 0.4 (0.0-1.0)  | 0.5 (0.0-2.2)   | -            | -                  | 0.23    |
| Monocytes, x10 <sup>9</sup> /L                                | -            | -            | -              | -              | -              | -               | -            | -                  | -       |

**Table 4.1: Clinical characteristics of CITE-seq cohort of n = 124 individuals.** Healthy volunteer (HV) and COVID-19 cohorts identical to chapter 3. Kruskal-Wallis tests performed for continuous data and chi-square test performed for count data. Data are n (%) or mean (SD) unless otherwise specified.



**Figure 4.1: Experimental design.** Gridplot of experimental design, depicting 10 pools (gPlex) in the columns, and 14 samples per pool (rows). Sample included per pool colored by comparator group status, with age and sex denoted by size and shape respectively.

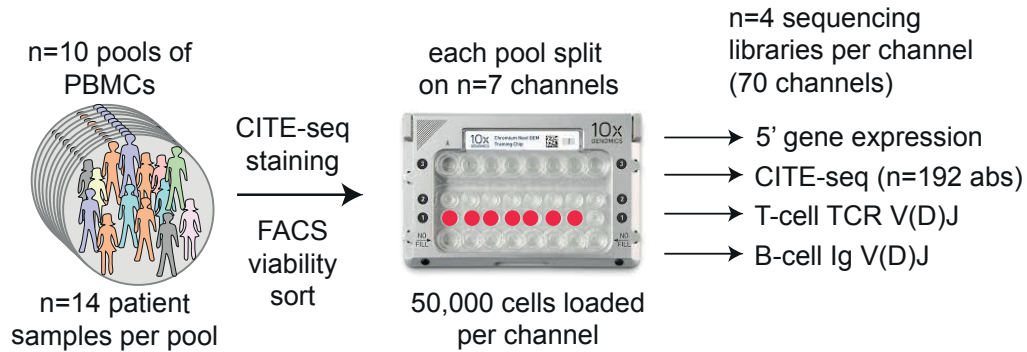
#### 4.2.2 CITE-seq preprocessing and quality control (QC)

After sequencing, preprocessing steps including read alignment, cell calling, genetic demultiplexing, and gene and cell filtering were performed (Chapter 2). After pre-processing, a total of 836,148 cells were retained for downstream analysis. Six samples had <1000 cells, including one sample from a critically ill COVID-19 patient (13 cells), one from a community mild COVID-19 case (23 cells) and four London flu samples (173, 637, 892 and 990 cells) (Fig. 4.3)

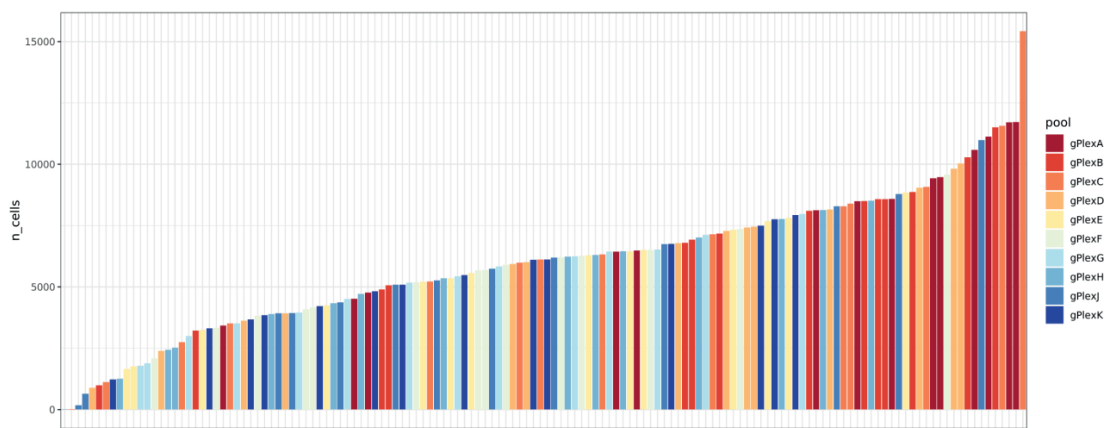
#### 4.2.3 Cell type/state clustering strategy

In order to best leverage the multi-modal nature of the experiment, and recognizing the difficulty of droplet based scRNA-seq methods in separating

### CITE-seq: Experiment overview



**Figure 4.2: Experimental workflow.** Experimental overview of n=140 samples split across 10 pools, with cell surface antibody staining (CITE-seq staining) and fluorescence activated cell sorting (FACS) for live cells, followed by single cell capture on the 10X platform where each pool of samples was loaded across 7 channels for sufficient cell capture. Each channel was then fed downstream into sequencing library preparation for gene expression, cell surface antibodies (CITE-seq) and T and B cell repertoires.



**Figure 4.3: Cell numbers.** Cell numbers per sample (columns) after QC for downstream clustering and differential analysis.

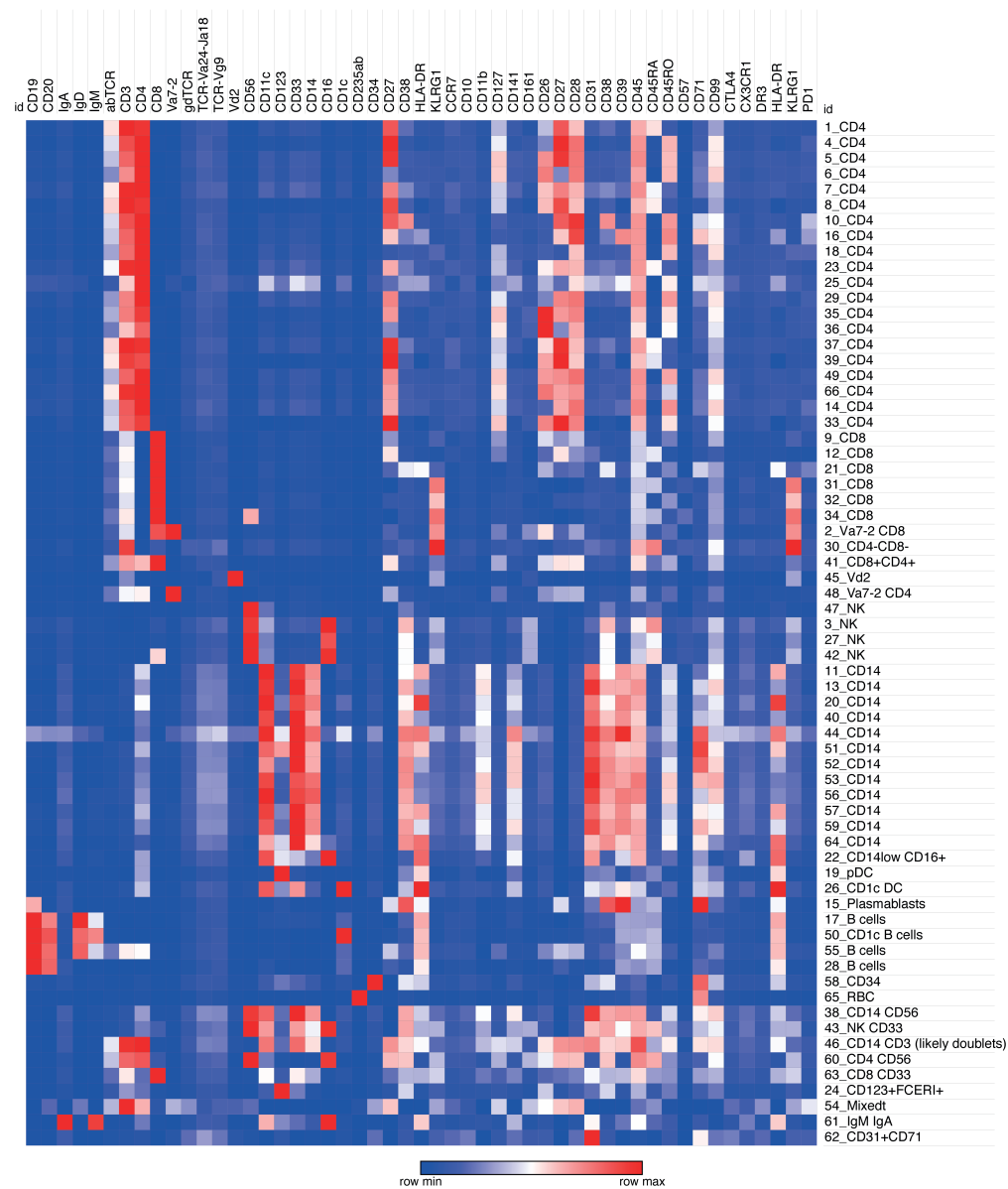
CD4+ vs. CD8+ T cell lineages apart (Hao et al. 2021), the cells were clustered with either GEX (graph based clustering) or ADT (unsupervised Gaussian mean shift clustering with CytoSplore (Höller et al. 2016)) features separately. The ADT clustering was prioritized for lineage definition. Using specific lineage markers within the 192 antibody panel (CD3, CD4, CD8, CD56, CD19, CD20, CD27, CD38, CD14, CD123, CD1c, CD33, CD235, CD34, V $\delta$ 2, V $\alpha$ 72 and CD161), the key lineages of mononuclear cells were identified including monocytes, NK cells, B cells, plasmablasts, plasmacytoid DCs, conventional DCs, red blood cells (RBC), CD34+ cells V $\delta$ 2+ gdT cells and CD3+ T cells (Fig. 4.4).

Based on this ADT prioritized clustering, each lineage was then further subjected to iterative GEX based clustering, and in the case of B and T cells, additional information from the repertoire feature space (not discussed herein) was included to refine annotations. Each lineage was clustered first to remove doublets, before a final round of clustering for detailed annotation. This resulted in three tiers of annotations: the broadest lineage/major subset definitions (CD4+ or CD8+ T cells, NK cells, mononuclear phagocytes [MNPs] divided into different monocyte and dendritic cells, B cells, plasmablasts and others), followed by minor subsets and 343 singlet immune clusters (detailed below) (COMBAT consortium 2022).

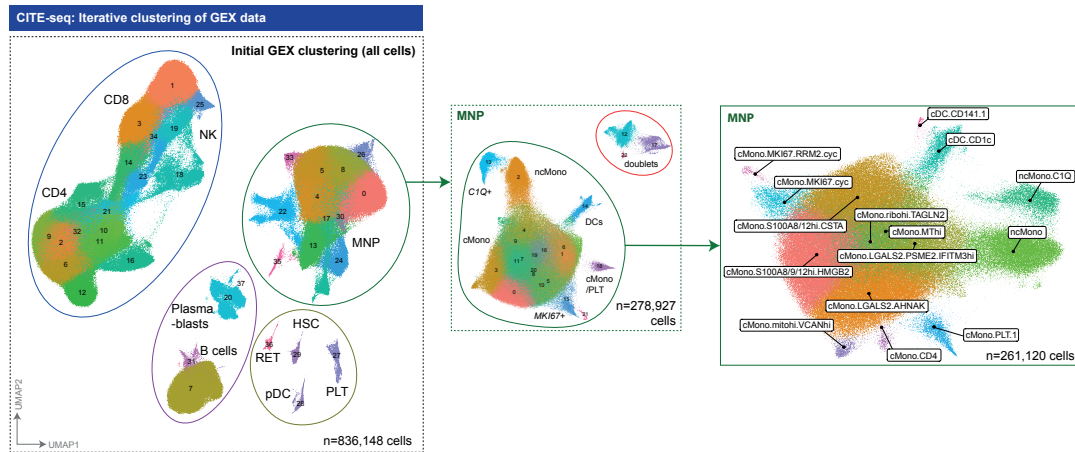
#### 4.2.4 Mononuclear phagocyte annotation

Following GEX based clustering of the MNP compartment, I identified well-described cell types including classical monocytes (cMono), non-classical monocytes (ncMono), conventional and plasmacytoid dendritic cells (cDCs and pDCs), as well as cMono with platelet contamination and cycling cMonos (Fig. 4.6). Surface protein expression of CD14/16 clearly separated cMonos from ncMonos (Fig. 4.7).

Detailed inspection of clustering revealed even further heterogeneity in the MNP



**Figure 4.4: Antibody derived tag (ADT) clustering for PBMC lineage identification.** PBMC lineage definitions by ADT clustering. Columns: selected surface proteins, rows: cluster assignments.



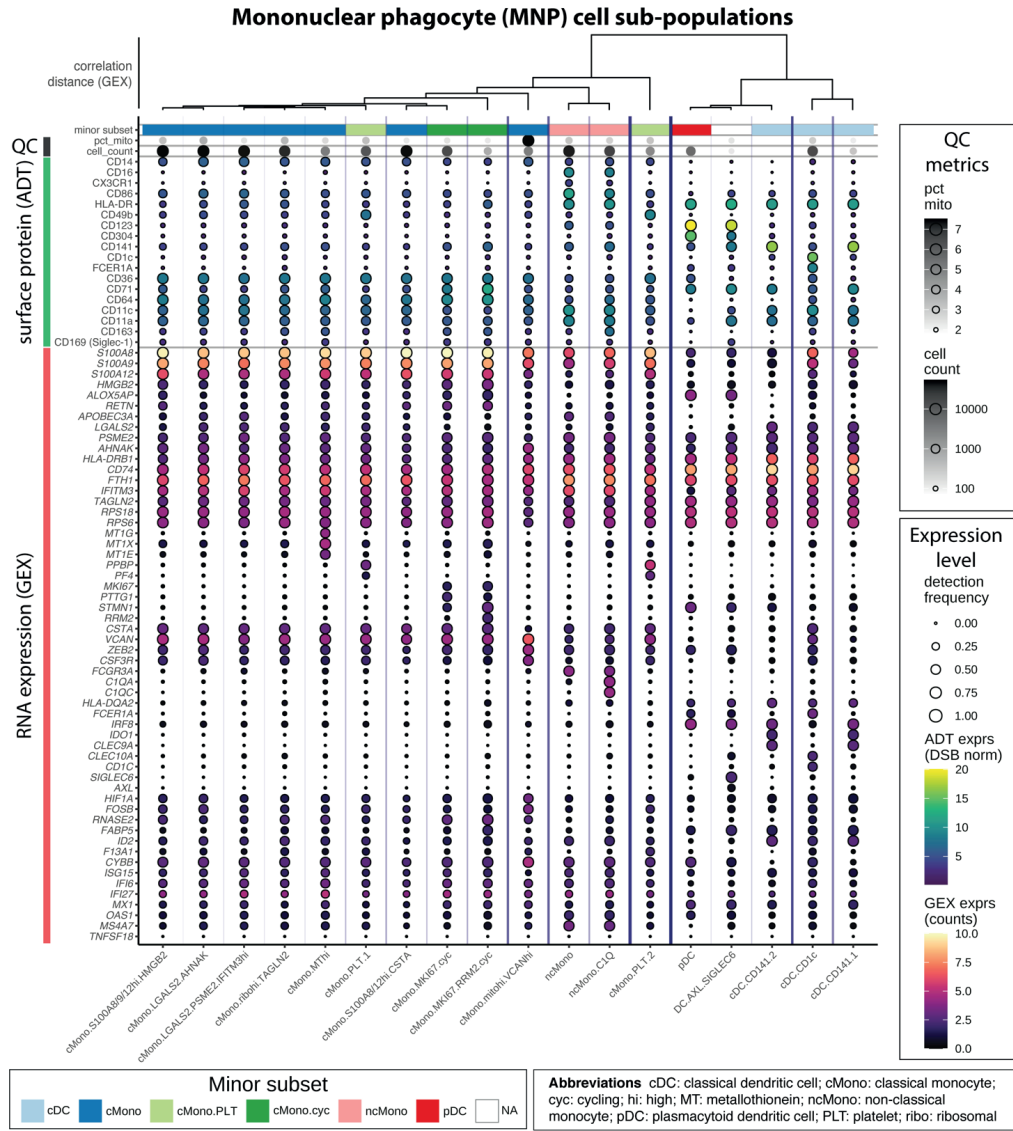
**Figure 4.5: Mononuclear phagocyte (MNP) subsetting and clustering strategy.** MNP subsetting from all cells after antibody derived tag clustering for lineage identification, followed by iterative gene expression (GEX) based clustering before detailed annotations.

compartment in the individual immune clusters. Conventional DC populations could be divided into CD141 positive cDCs and CD1c positive cDCs, which could be seen in both GEX and ADT, in addition to AXL/SIGLEC6 (GEX) positive cDCs, recapitulating traditional understanding of DC subsets as well as more recent definitions of DC heterogeneity (Villani et al. 2017). In turn, ncMonos were separately clustered into C1Q+ vs. C1Q- subsets, and cMonos were annotated based on their expression of *S100A8/9*, *LGALS2*, *AHNAK*, *PSME2*, *TAGLN2*, *CSTA*, *VCAN*, mitochondrial genes, and *MKI67* and *RRM2* for cycling cMonos.

Given that the majority of the dataset consisted of samples from patients with a viral infection, I looked for whether there were distinct subsets of cells responsive to IFN, but genes including IFN responsiveness (*ISG15*, *IFI8*, *IFI27*, *MX1* and *OAS1*) showed a similar pattern of expression across most monocyte subset, suggesting against a specific group of IFN responsive MNPs (Fig. 4.6).

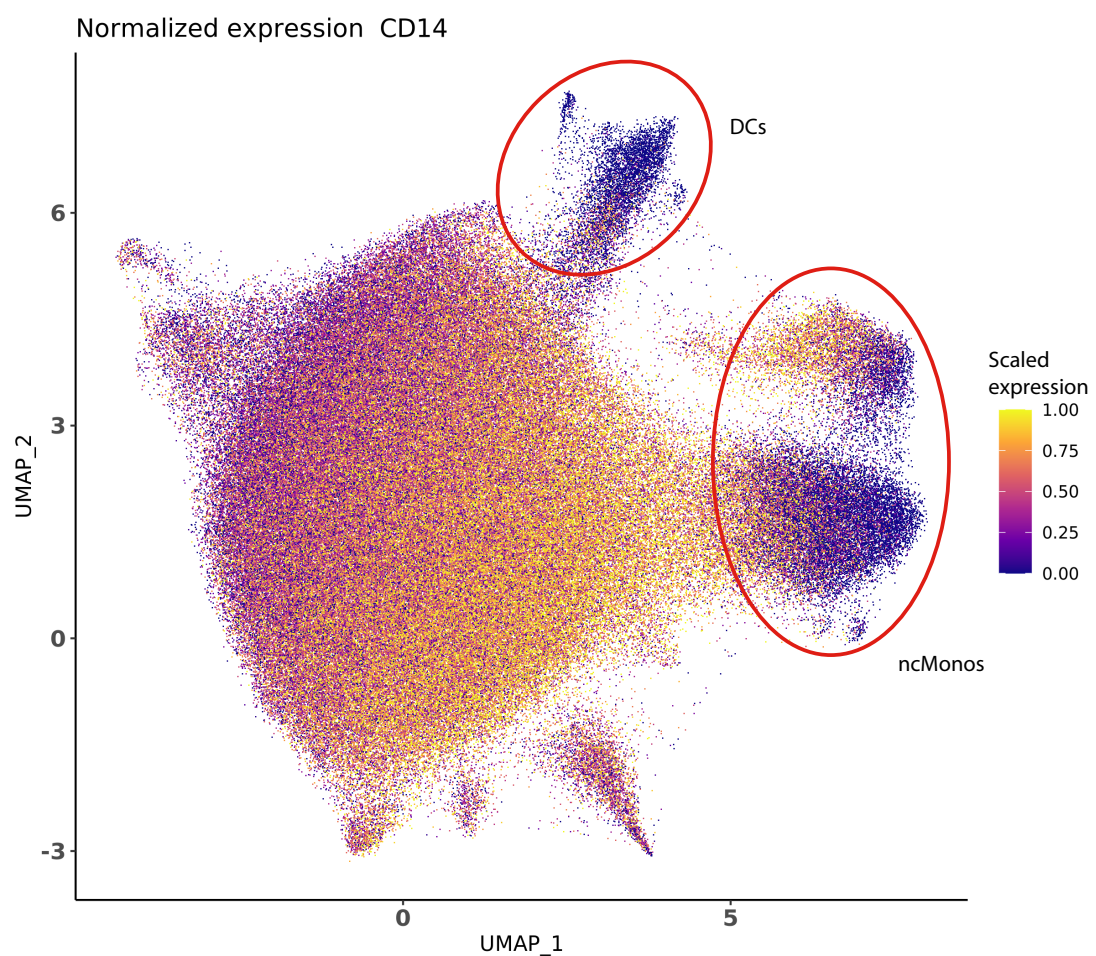
#### 4.2.5 T/NK cell annotation

CD4+ and CD8+ T cells were clustered and annotated according to existing immunological understanding of naive, memory and effector cell states, with cell surface CD45RO and CD45RA expression and the naive gene markers



**Figure 4.6: Mononuclear phagocyte (MNP) annotations.** MNP subsets annotated. Marker surface proteins (top, green rows) and genes (bottom, red rows) used for annotations are shown. Quality control (QC) metrics are also shown at the top in terms of cells per cluster (cell count) and percentage of mitochondrial genes detected per cluster (pct mito). ADT = antibody derived tags, exprs = expression, DSB = denoised and scaled by background, GEX = gene expression





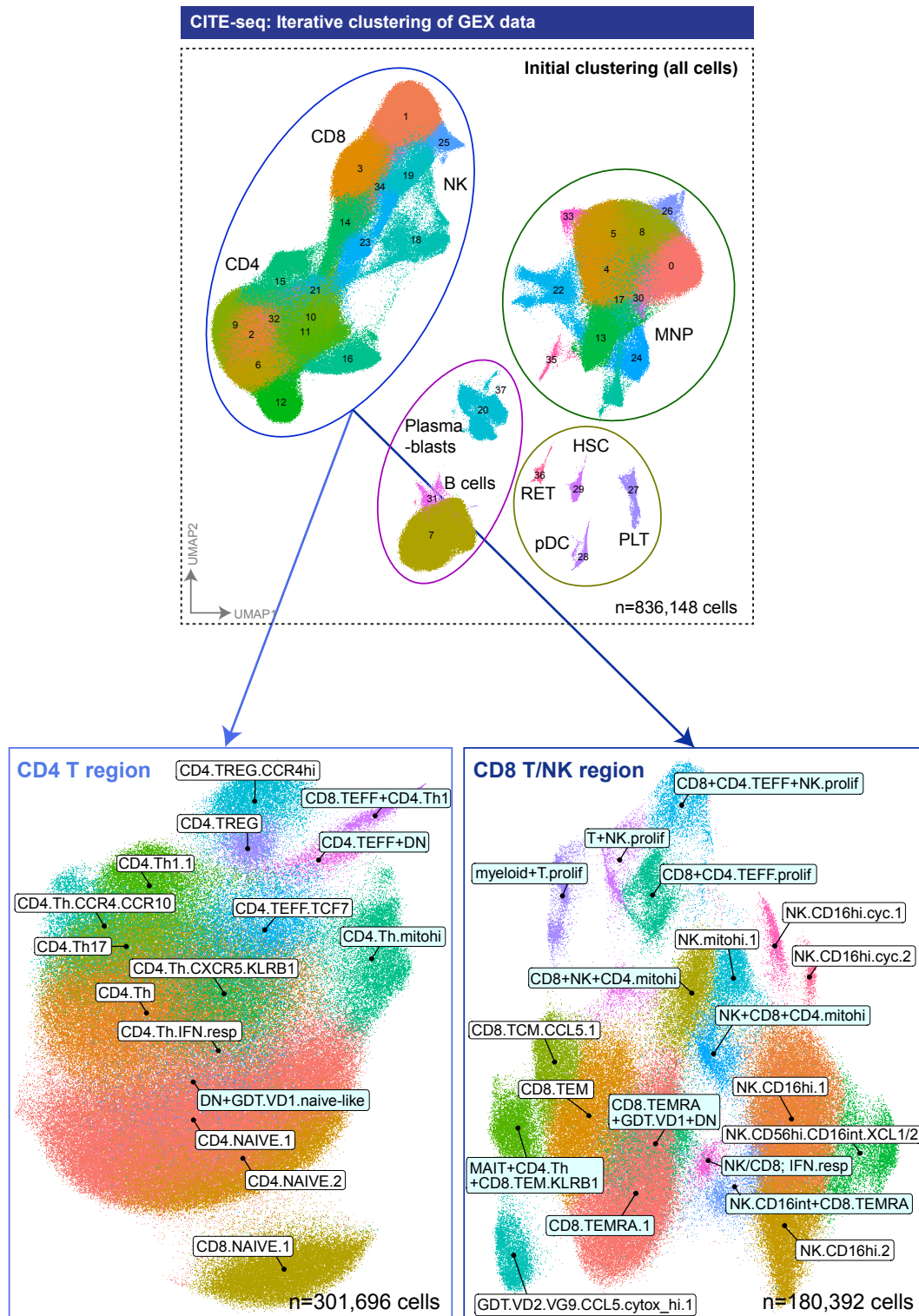
**Figure 4.7: Mononuclear phagocyte (MNP) CD14 expression.** MNP denoised and scaled by background cell surface CD14 expression overlaid on MNP UMAP embeddings for delineation of classical monocytes vs. non-classical monocytes and dendritic cells.



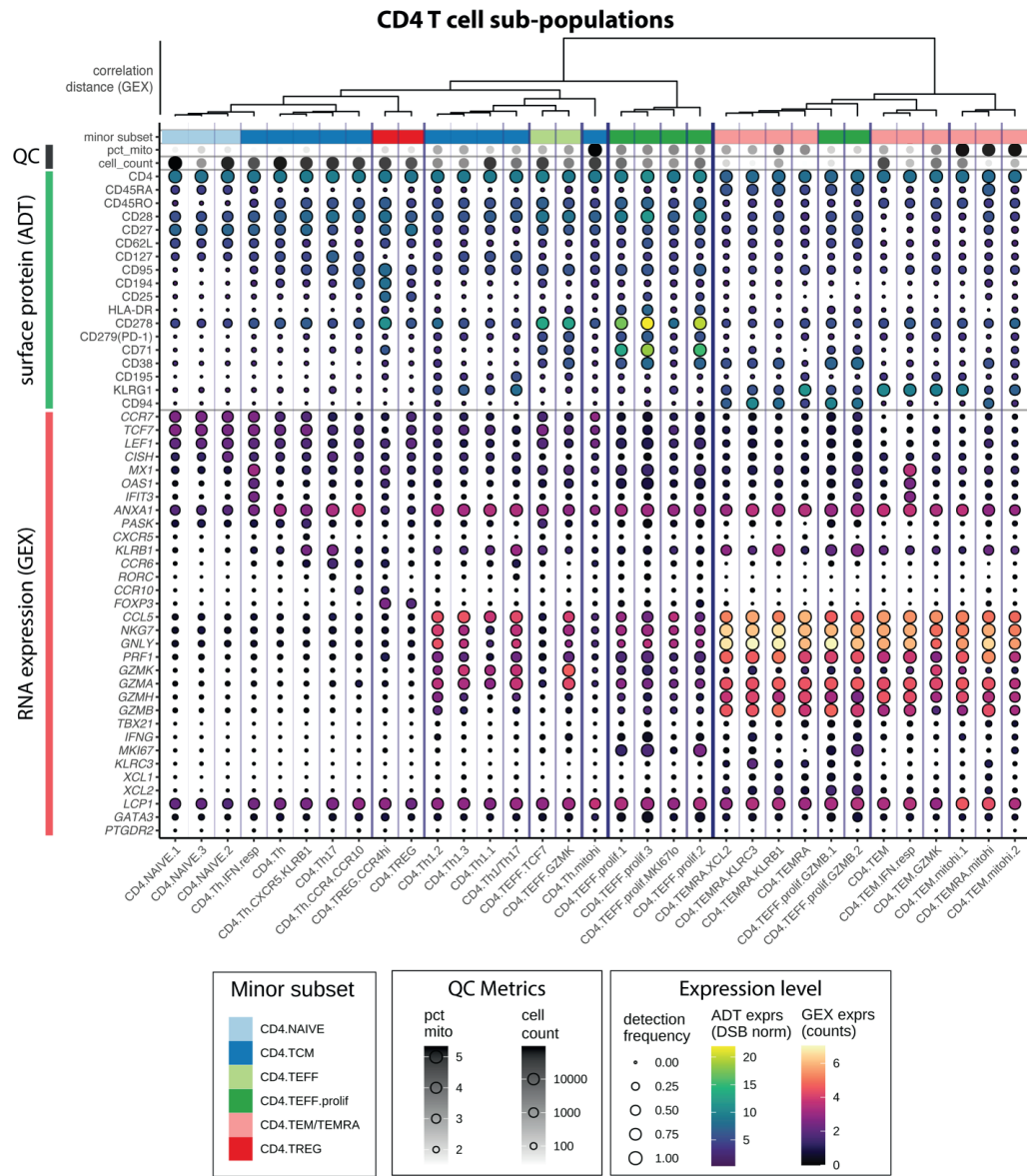
*LEF1/TCF7/CCR7* being the key markers for splitting between naive vs. memory populations.

For the CD4s, the effector memory cells clusters (TEM/TEMRA) clustered together while the naive, central memory (TCM), regulatory T cells (Tregs), and effector T cells (TEFF) clustered more closely with each other in gene expression space, while proliferating TEFF clusters were split between both groups (Fig. 4.9). Notably, I observed a gradient of effectorness moving from more naive T cell states to more effector/TEM/TEMRA states as measured by expression of genes related to cytotoxic function such as *CCL5*, *GNLY*, *PRF1*, *GZMA*, *GZMB*, *GZMH* (Cano-Gamez et al. 2020). Cycling cells were marked out by their *MKI67* expression, similar to the MNPs. Our fine annotations of the CD4+ cell states further revealed three clusters of TCMs with a Th1 phenotype given *IFNG* and *TBX21* expression, whereas TCM populations with more *CCR6* expression were labelled Th17. Notably, I could not identify Th2 subsets through this multimodal data, suggesting against hypotheses that Th2 cell expansion led to poor outcomes in infection (either severe/septic COVID-19 or non-COVID-19 sepsis). TEFF were given the designation based on a combination of high cell surface expression of CD278 (ICOS), CD38, HLA-DR, PD-1, and were split into two clusters, one expressing more GEX markers similar to naive cells (*CCR7/TCF7/LEF1*) while the other expressed more genes of cytotoxic effector function (*CCL5/GZMK/GZMA*).

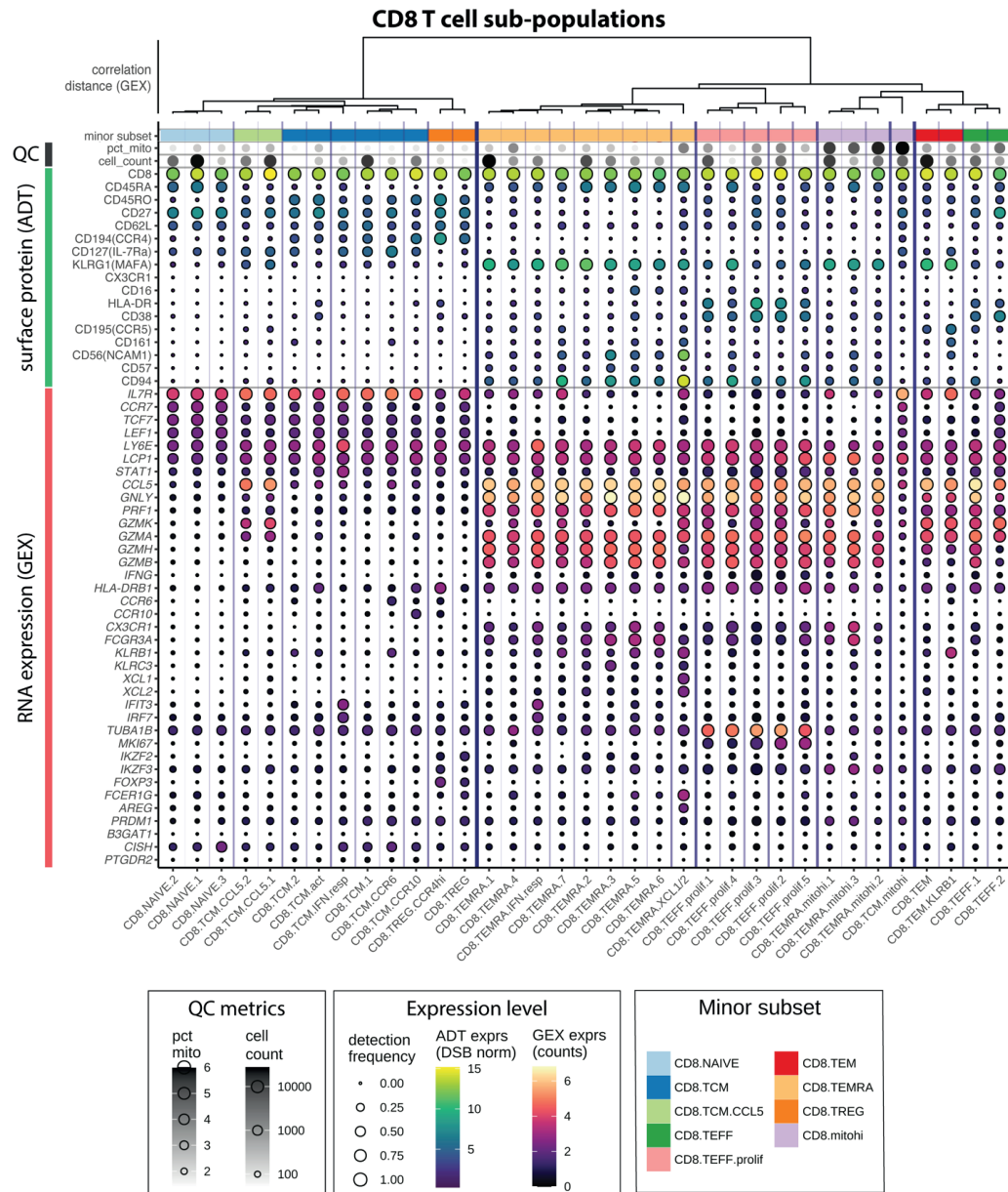
Similarly for CD8s, two broad groups could be seen on hierarchical clustering, including naive, TCM and CD8+ Tregs in one group and TEM, TEMRA, proliferating and non-proliferating TEFF, as well as cells high in mitochondrial genes (Fig. 4.10) in another group. Proliferating cells were clearly distinguished by their *MKI67* expression. Naive cells exhibited higher cell surface CD27 and CD62L expression along with *CCR7*, *TCF7*, and *LEF1* in the gene expression space in contrast to memory/effector cells. CD8+ Tregs were found in this dataset based on their *FOXP3* expression, but were very few in number. TEM



**Figure 4.8: T and natural killer (NK) cell subsetting and clustering strategy.** T and NK cell subsetting from all cells after antibody derived tag clustering for lineage identification, followed by iterative gene expression (GEX) based clustering with CD4+ T cells split apart from CD8+ T cells and NK cells for detailed annotations.



**Figure 4.9: CD4+ T cell annotations.** CD4+ T cell subsets annotated. Marker surface proteins (top, green rows) and genes (bottom, red rows) used for annotations are shown. Quality control (QC) metrics are also shown at the top in terms of cells per cluster (cell count) and percentage of mitochondrial genes detected per cluster (pct mito). ADT = antibody derived tags, exprs = expression, DSB = denoised and scaled by background, GEX = gene expression



**Abbreviations** act: activated; hi: high; IFN: interferon; lo: low; mito: mitochondrial; prolif: proliferating; resp: responsive; TCM: T central memory; TEFF: T effector; TEM: T effector memory; TEMRA: CD45RA+ T effector memory; Th: T-helper; TREG: regulatory T cell. **Symbols**: x: NA.

**Figure 4.10: CD8+ T cell annotations.** CD8+ T cell subsets annotated. Marker surface proteins (top, green rows) and genes (bottom, red rows) used for annotations are shown. Quality control (QC) metrics are also shown at the top in terms of cells per cluster (cell count) and percentage of mitochondrial genes detected per cluster (pct mito). ADT = antibody derived tags, exprs = expression, DSB = denoised and scaled by background, GEX = gene expression



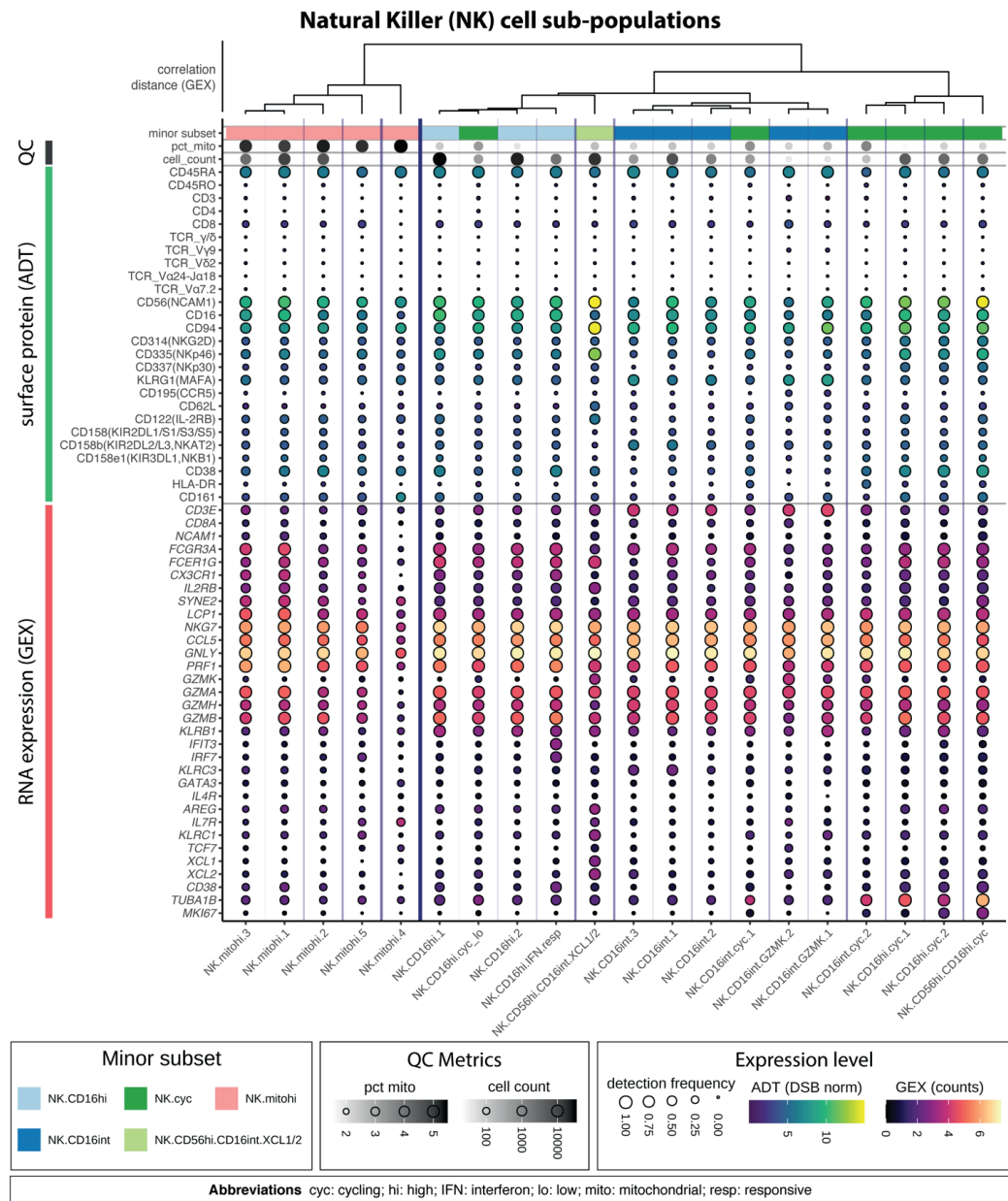
vs. TEMRA distinctions were made based on cell surface CD45RA expression to align with existing understanding of CD45RA-expressing TEM cells which are known to exhibit particularly strong cytotoxic properties. Indeed, inspection of gene expression signatures showed that TEMRA cells had higher *GZMH*, *GZMB* and *PRF1* expression than TEM cells, supporting the notion that they possess a potent cytotoxic phenotype.

For NK cells, I found a number of clusters expressing a relatively high percentage of mitochondrial genes which clustered separately to the other populations, and again cells which expressed gene markers of cell cycling (*MKI67*, *TUBA1B*) (Fig. 4.11). I found in contradistinction to existing literature classically dividing NK cells into CD56 bright vs. dim NK cells (Marcenaro et al. 2017) that most of the NK cells in this dataset expressed CD56 on the cell surface at a similar level except for two specific clusters with particular high CD56 expression, which I therefore specifically appended the level of CD56 bright to. For the remaining cells, I therefore designated the clusters as expressing CD16 (*FCGR3A* gene) either at a high or intermediate level.

Whereas there was no clear cluster of IFN responsive MNPs, I observed distinct IFN responsive clusters in all of the CD4/8 T cell and NK cell compartments (Fig. 4.9, 4.10, 4.11), marked out by common *IFIT3* expression. In the case of CD4+ T cells, *MX1* and *OAS1* were also expressed, while in CD8+ T cells and NK cells *IRF7* was expressed.

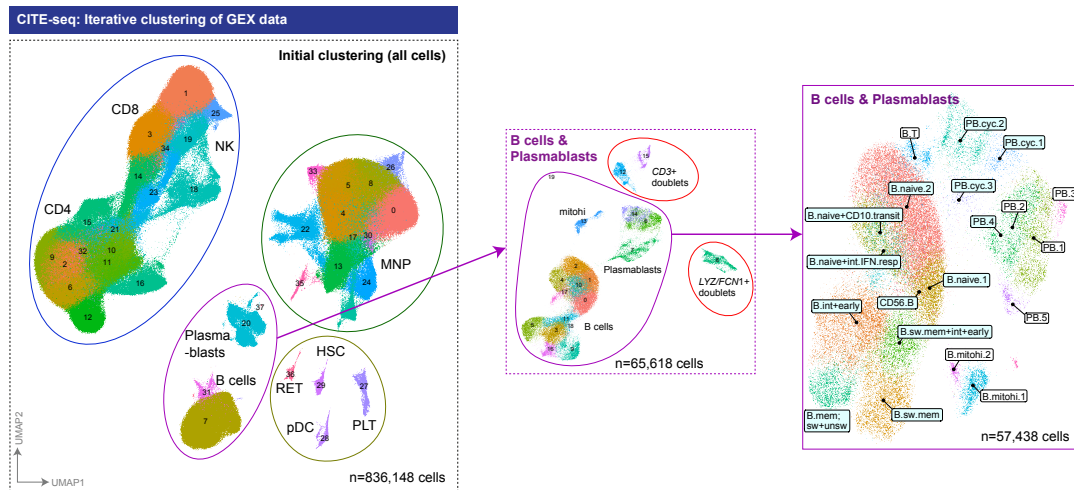
#### **4.2.6 B cell annotation**

The B cell compartment clearly separated into plasmablasts (low surface CD20 expression) and B cells. Within the B cells, I annotated naive, intermediate and memory populations based on surface IgD/M expression, with cells expressing IgD/M denoting naive or intermediate cells (Fig. 4.13). I also used surface CD27 and CD38 to further distinguish between memory (CD27+), naive (CD27-) and



**Figure 4.11: Natural killer (NK) cell annotations.** NK cell subsets annotated. Marker surface proteins (top, green rows) and genes (bottom, red rows) used for annotations are shown. Quality control (QC) metrics are also shown at the top in terms of cells per cluster (cell count) and percentage of mitochondrial genes detected per cluster (pct mito). ADT = antibody derived tags, exprs = expression, DSB = denoised and scaled by background, GEX = gene expression.

intermediate (CD27 intermediate) subsets, allowing us to identify a cluster of unswitched memory B cells (CD27+ CD38- IgM/D+). Our extensive ADT panel allowed us to annotate rarer B cell subsets including CD1C positive B cells, as well as CD10 transitional B cells which represent B cells recently emigrant from the bone marrow (Palanichamy et al. 2009).

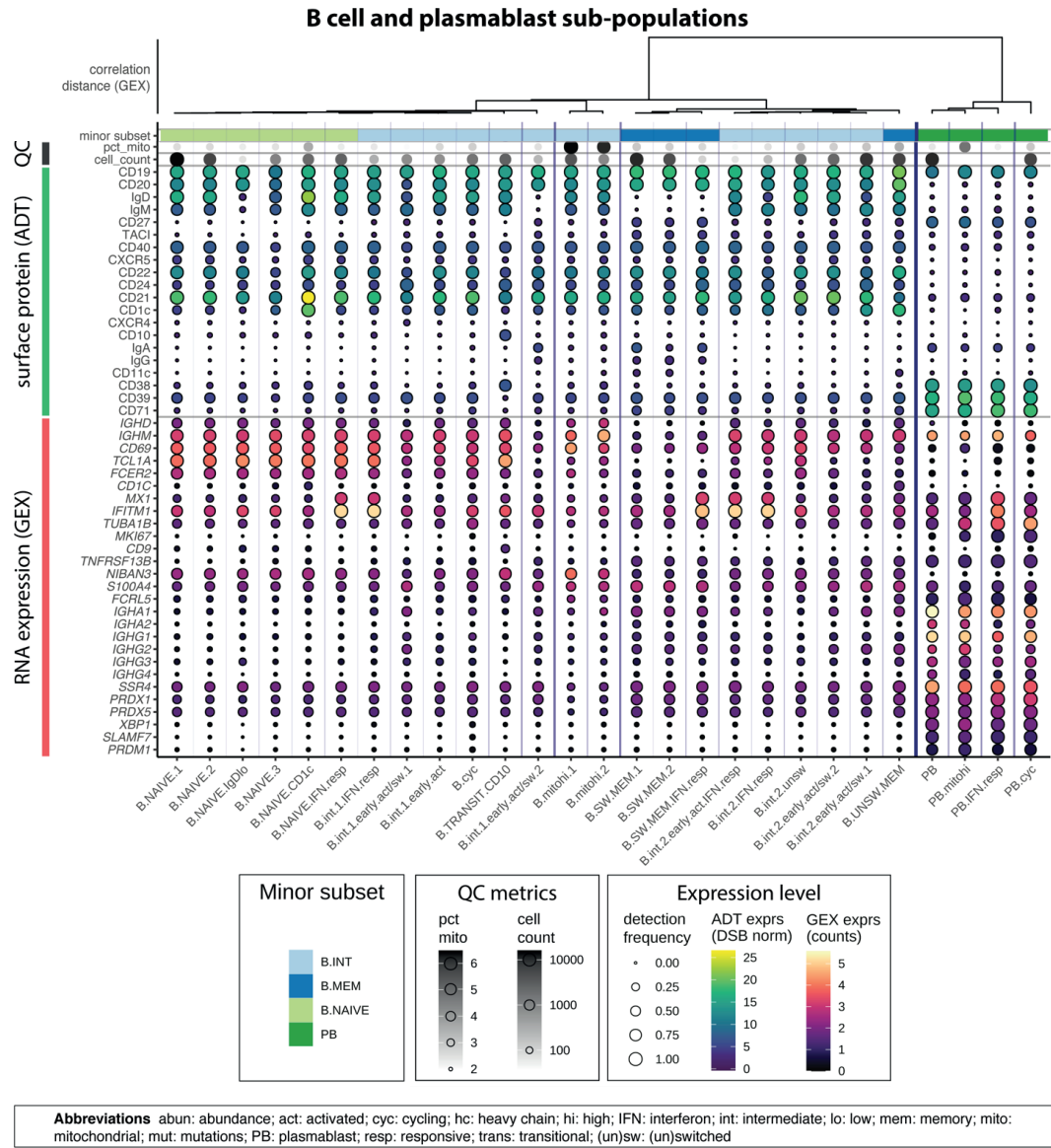


**Figure 4.12: B cell and plasmablast subsetting and clustering strategy.** B cell and plasmablast subsetting from all cells after antibody derived tag clustering for lineage identification, followed by iterative gene expression (GEX) based clustering for detailed annotations.

The early activation label was given to B cell clusters where BCR analysis by collaborators (Bashford-Rogers group) showed either usage of *IGHA*, *IGHE* or *IGHG*, or if there were  $\geq 2$  mutations in the *IGHD/IGHM* genes (COMBAT consortium 2022). Similar to the T/NK cells, I observed clusters of B cells with IFN stimulated gene expression (*IFITM1*, *MX1*) (Fig. 4.13).

#### 4.2.7 Other cell annotation

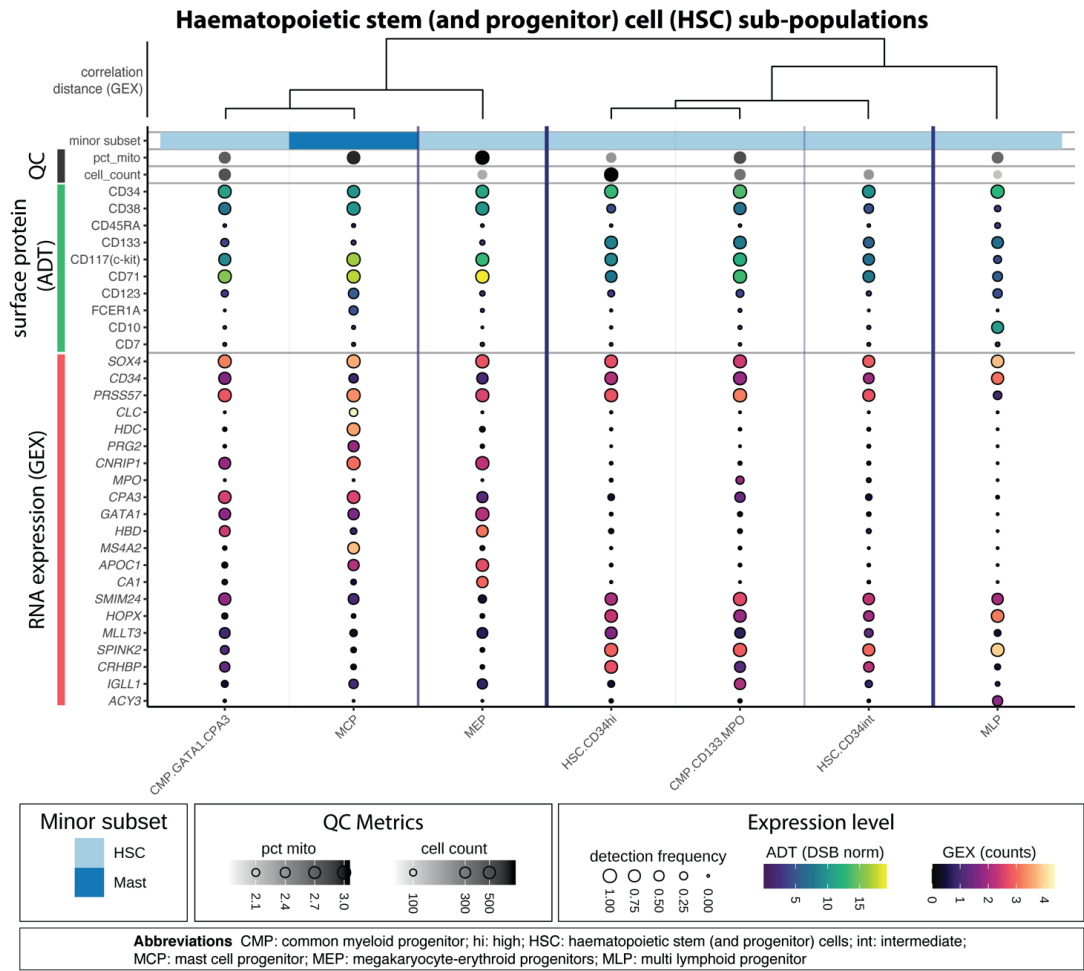
Aside from the major lineages of B, T, NK cells and MNPs, I was also able to annotate a number of hematopoietic stem and progenitor cell (HSPC) clusters (Fig. 4.14). I identified CD34 high HSCs, as well as two clusters of common myeloid progenitors (expressing *MPO* and *GATA1/CPA3* respectively). On the basis of surface CD10 expression, I further annotated a cluster of HSPCs



**Figure 4.13: B cell and plasmablast annotations.** B cell and plasmablast subsets annotated. Marker surface proteins (top, green rows) and genes (bottom, red rows) used for annotations are shown. Quality control (QC) metrics are also shown at the top in terms of cells per cluster (cell count) and percentage of mitochondrial genes detected per cluster (pct mito). ADT = antibody derived tags, exprs = expression, DSB = denoised and scaled by background, GEX = gene expression.



as multipotent lymphoid progenitors (MLPs). Finally, I found a cluster of megakaryocyte erythroid progenitors with *HBD* expression.



**Figure 4.14: Hematopoietic stem and progenitor cell (HSPC) annotations.** HSPC subsets annotated. Marker surface proteins (top, green rows) and genes (bottom, red rows) used for annotations are shown. Quality control (QC) metrics are also shown at the top in terms of cells per cluster (cell count) and percentage of mitochondrial genes detected per cluster (pct mito). ADT = antibody derived tags, exprs = expression, DSB = denoised and scaled by background, GEX = gene expression.

In summary, the CITE-seq atlas permitted successful identification of PBMC subpopulations encompassing a range of previously described cell types and states within a single experiment using a unified profiling platform rather than relying on scattered definitions deployed in each individual experiment.

#### 4.2.8 Differential analysis

In addition to the atlasing bioresource, one of the aims was to understand the immune remodeling that defined all-cause sepsis / COVID-19 viral sepsis. To tackle this question, I performed two sets of differential analyses, first looking at differential abundance of the various cell types/states identified from the above multimodal annotation, and second looking at pseudobulked cell type weighted gene correlation network analysis to identify coordinated changes in gene expression patterns across conditions.

#### 4.2.9 Cell type/state differential abundance analysis

##### Minor subset differential abundance analysis

Differential abundance (DA) analysis was performed at multiple levels, first at the level of gross MNC cell types ("minor subsets"), then within each MNC compartment ("individual immune clusters").

At the level of minor subsets, I found contrasting the various comparator groups with HVs that in severe and critical COVID-19, the cell types differentially more abundant included cycling populations such as cMono and effector CD4<sup>+</sup> T cells, as well as plasmablasts and platelets/CD34 megakaryocyte progenitors and hematopoietic stem cells (Fig. 4.15A; Appendix). Contrasting sepsis samples to HVs showed a similar result except for a lack of plasmablast enrichment in sepsis. In addition, whereas cycling effector CD4<sup>+</sup> T cells were enriched in CC/CS and flu/sepsis, it was the proliferating effector CD8<sup>+</sup> T cells that were higher in CM.

To link these changes to potential functional implications, I tested for association of abundance of each of the PBMC minor subsets with various clinical parameters of interest. From observations of the clinicians within the COMBAT consortium, specific COVID-19 phenotypes beyond standard clinical/laboratory measures of severity which were deemed relevant included persistent fever or fever at

sampling, and whether the patient suffered from a radiological thromboembolus.

While no particular cell type was associated with survival, I found that the platelet/CD34 megakaryocyte progenitors, cycling cMono and HSC were associated with a number of parameters linked to severity of disease including World Health Organisation COVID-19 ordinal status and ventilation/SOFA O<sub>2</sub>/oxygen status at sampling (Fig. 4.15B).

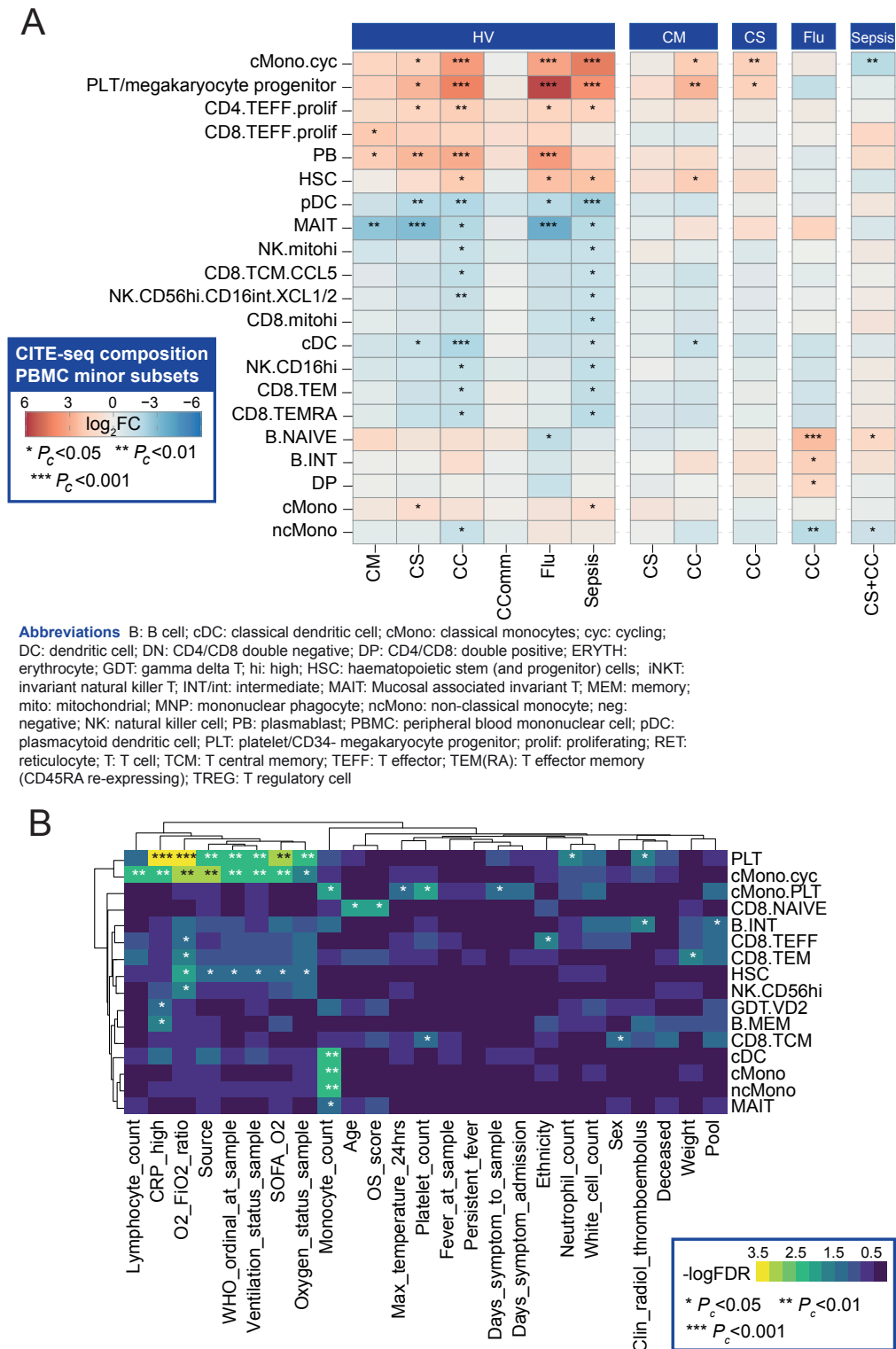
Regarding specific COVID-19 phenotypes, there was no statistically significant relationship between the cell types and fever status, but notably there was a significant association between platelet/CD34 megakaryocyte progenitor abundance with thromboembolism in hospitalized COVID-19 patients (Fig. 4.15B).

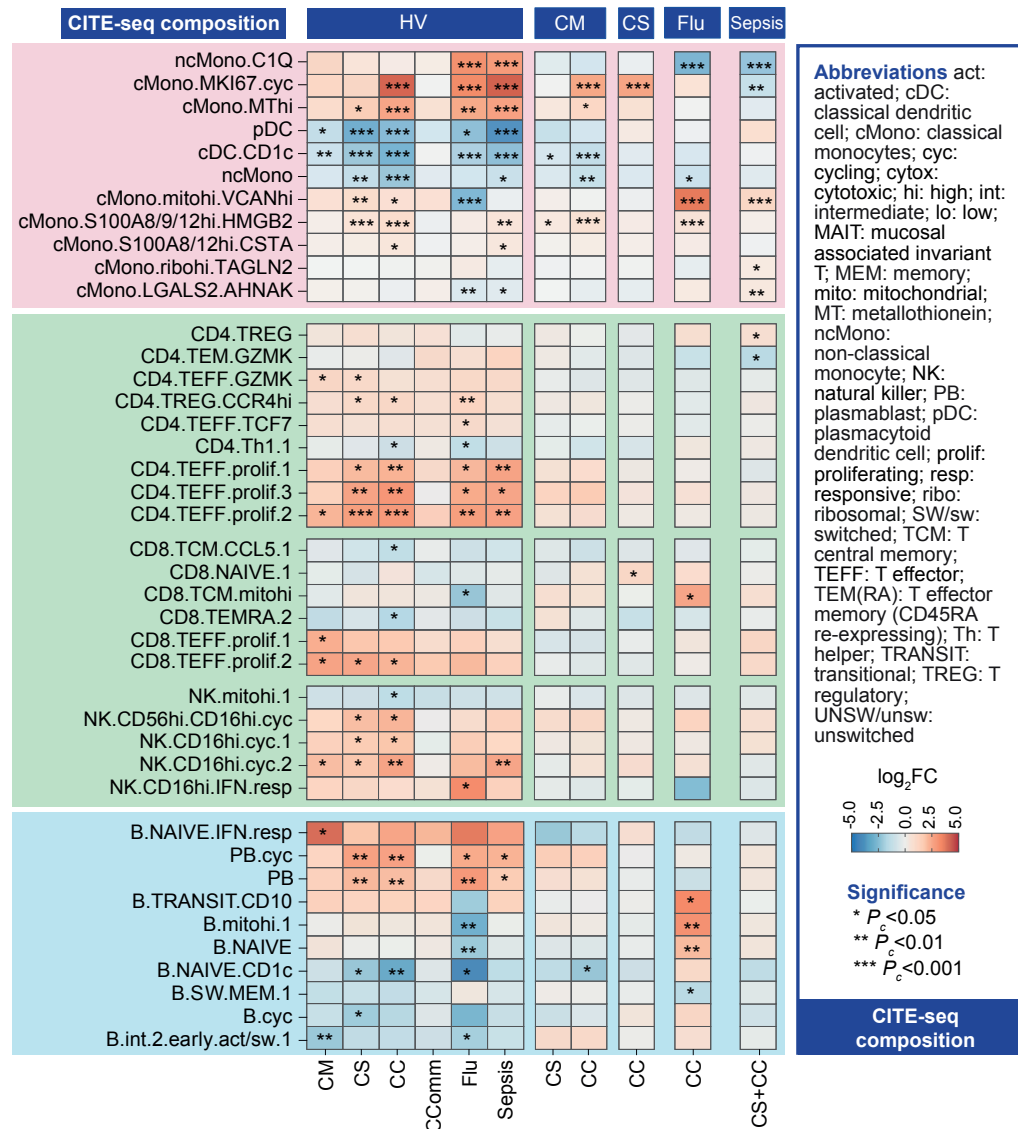
The signature of cell types reduced in either critical COVID-19 or sepsis vs. HV was similar, with reductions in pDCs, cDCs, MAIT cells, mitochondrial high NK cells, CD56hi CD16int *XCL1*/2+ NK cells, CD16hi NK cells, *CCL5*+ CD8+ TCMs and CD8+ TEMs, with mitochondrial high CD8+ T cells and CD8+ TEMRAs additionally also less abundant in sepsis, suggesting an even greater depletion of cytotoxic (CD8+ T/NK) cells. However, when directly contrasting sepsis samples with critical COVID-19, the differential abundance changes were much milder, with a decrease in ncMono and cDCs seen in critical COVID-19 and a slightly greater proportion of naive B cells. There was no significant change found across the T cell compartment. This was likewise the case when comparing either sepsis with COVID-19 severe, or COVID-19 critical to severe.

### **Detailed cell type/state differential abundance analysis**

To understand the changes within specific lineages in more detail, I performed differential abundance analysis for CD4+ T cells, CD8+ T cells and NK cells, and MNPs at the fine annotation level of individual immune clusters.

For the MNPs, contrasts with HVs revealed a depletion of DCs with increasing





**Figure 4.16: Composition analysis of PBMC high resolution immune clusters across conditions.** Differential abundance analysis for patients vs. healthy volunteers and across disease conditions with empirical Bayes methodology, with clusters displayed where there is a significant difference in at least one contrast. Cluster frequency calculated as a total of total myeloid cells, total B and plasmablast cells, or total T and natural killer (NK) cells.

COVID-19 severity and in flu/sepsis, as well as reduced ncMonos in CS/CC and sepsis (Fig. 4.16). The greatest relative increases in abundance compared to HVs were seen in *MKI67*+ cycling cMonos and Metallothionein hi cMonos for flu/CC/sepsis as well as CS for the latter, and *C1Q*+ ncMonos for flu and sepsis. To confirm the specificity of *C1Q*+ ncMonos in distinguishing between flu/sepsis against COVID-19, the contrasts between flu or sepsis vs. CS/CC indeed showed depletion of *C1Q*+ ncMonos in CS/CC (Fig. 4.16). By contrast, the *VCAN*hi mitochondrial hi cMonos appeared to be specifically enriched in COVID-19, with contrasts for CS/CC against flu/sepsis showing increased abundance of the cluster. Other clusters of monocytes with changes across conditions included the cMonos expressing *S100A8/9* genes, with the *HMGB2* expressing cluster also elevated in CC vs. flu, but not when comparing against sepsis.

Detailed analysis of CD4+ T cells recapitulated the increases in proliferating effector T cells in severe/critical COVID-19 vs. HVs, with one of the proliferating effector CD4+ T cell clusters (cluster 2) even showing enrichment in CM vs. HV (Fig. 4.16). However this analysis did not reveal particularly striking features separating COVID-19, flu or sepsis except a mild change in Tregs in CS+CC vs. sepsis and CD4+ TEMs (Fig. 4.16). When looking at the CD8+ T cells / NK cells, I found a higher proportion of proliferating CD8+ effector T cells (cluster 2) that was more abundant across all COVID-19 severities compared to HV while one of the clusters (cluster 1) was only more abundant in CM, contrasting slightly with the findings in gross MNC analysis where proliferating effector CD8+ T cells were higher than in HVs only in CM (Fig. 4.16). Within the NK cell compartment, the changes occurred mostly in terms of increases in cycling CD16hi NK cell clusters in CS/CC vs. HV, but there were again no significant differences differentiating CS/CC vs. flu/sepsis. Whereas the IFN responsive cluster was enriched in flu vs. HVs, there was no significant increase in the abundance of those cells in any of the COVID-19 groups.

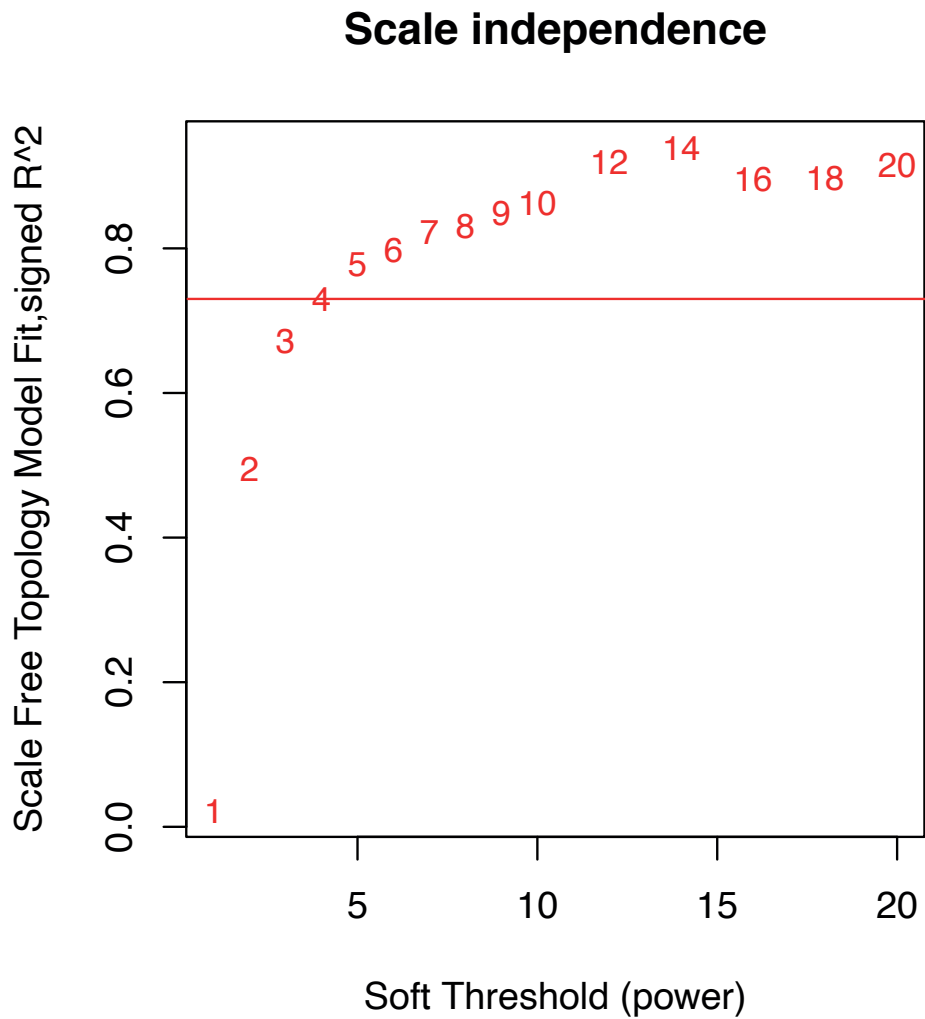
In the B cell compartment, the increase in PBs in the disease conditions vs. HV was recapitulated (Fig. 4.16). When contrasting the various disease conditions with each other, the key result was that there was an enrichment of various B cell clusters including naive and transitional CD10+ B cells and depletion of switched memory B cells in CC vs. flu.

#### 4.2.10 Pseudobulk cell type WGCNA

One limitation to DA analysis is the difficulty of understanding the function of specific cell clusters as called by clustering algorithms. To complement the DA analysis, I therefore examined the gene expression changes of the different cell types across the various disease states to infer changes in cell function. I did this by pseudobulking each sample by major cell subset i.e. cMonos, ncMonos, CD4+ T cells, CD8+ T cells, NK cells, B cells and PB, and then conducting WGCNA for each individual cell type.

In brief, the analysis for each cell type was as follows. First, I took the log2-transformed counts-per-million data per sample and built a signed-hybrid gene co-expression network with biweight midcorrelation. A signed network was chosen as the direction of correlation likely has biological relevance, and the biweight midcorrelation implementation permitted robust correlation measures to be calculated in the setting of missing data (due to sparsity of scRNA-seq data). A soft thresholding power of four was used, as this reached a scale free topology model fit of  $>0.7$  for most cell types (Fig. 4.17). This adjacency matrix was then transformed into a topological overlap matrix to calculate the dissimilarity, and the genes were clustered into modules. Modules with eigengene (the first principal component of a module) dissimilarity  $<0.3$  were merged (Fig. 4.18). At this level of module merging, I found seven modules for cMono, ten for ncMono, seven for CD4+ T cells, eight for CD8+ T cells, nine for NK cells, 11 for B cells and three for PB respectively. The modules were then clustered and

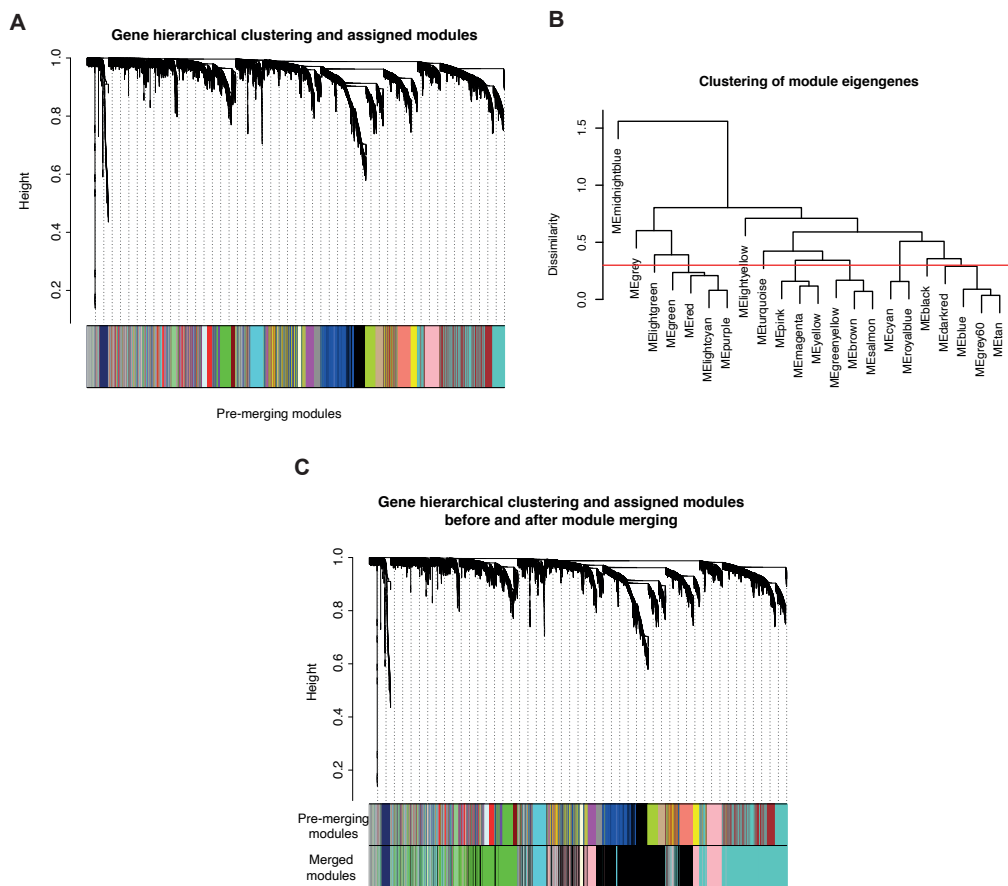
tested against various clinical parameters (Fig. 4.19A). Following discovery of cell type based modules, these modules were annotated by pathway enrichment analysis of the genes contained in the modules against multiple databases (GO biological functions, molecular processes and cellular components, as well as KEGG and Reactome databases) (Fig. 4.19B).



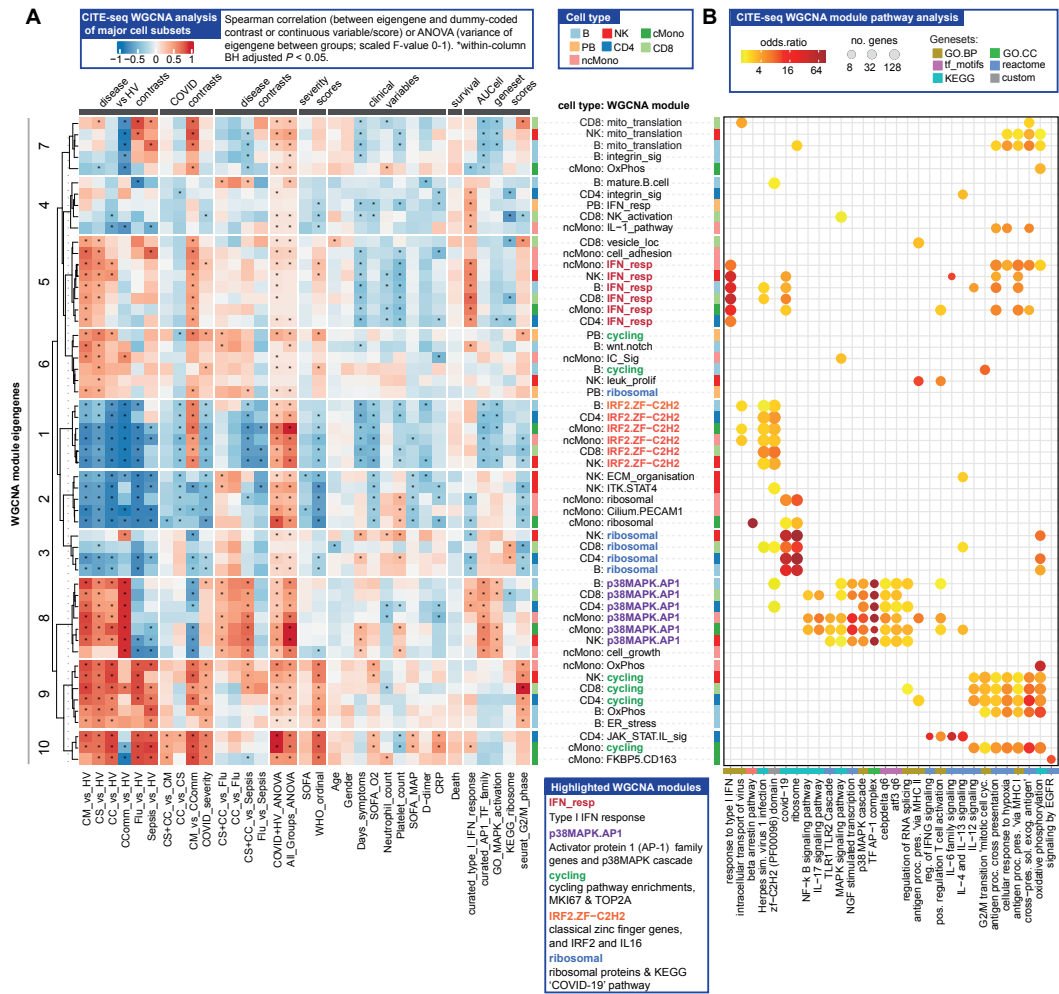
**Figure 4.17: Scale-free network soft thresholding decision for weighted adjacency matrix construction for weighted gene correlation network analysis (WGCNA).** Scale free topology model fit at varying soft thresholding levels for WGCNA. Representative example for CD4+ T cells depicted here.

From the annotation and clustering analysis of WGCNA modules, I identified five module sets out of the 55 modules which were shared to varying extents across





**Figure 4.18: Weighted gene correlation network analysis module construction and merging.** (A) Hierarchical clustering and module assignment of genes based on weighted adjacency matrix. (B) Clustering of module eigengenes with threshold of correlation dissimilarity of 0.3 for module merging shown by red line. (C) Gene module assignment after module merging.



**Figure 4.19: Weighted gene correlation network analysis (WGCNA) of gene expression modules for PBMC major cell types.** (A) Association of module eigengenes with disease contrasts, clinical features, survival and gene set scores (all significant associations shown). (B) Module pathway enrichment for testing against Gene Ontology (GO), Reactome and KEGG databases.

the cell types (Fig. 4.19A). First, I found a set of type I IFN response modules in all cell types except PB (i.e. six modules), including in the cMonos and ncMonos (where no distinct IFN cluster of cells was seen). Intriguingly, I found that while these modules were significantly higher in expression in CM vs. HVs, they were not higher in CC vs. HV and only four were significantly upregulated in CS vs. HV. A direct contrast of CM vs. CS+CC showed no significant difference.

To further understand the nature of the IFN response in COVID, I correlated the IFN module expression with other clinical variables (Fig. 4.19A), and found that the expression of this module was actually correlated with time of sampling from symptom onset, with a higher IFN response earlier in disease course. Given the relatively later sampling for CC samples, this was likely the explanation for the lack of a significant difference in IFN module expression between CC vs. HV.

The second module set, also discovered in all cell types except PB, showed strong enrichment for activator protein 1 (AP-1) (FOS, JUN, ATF family genes) and the p38MAPK cascade (p38MAPK.AP-1 module) (Fig. 4.19A-B). I confirmed the relevance of the AP-1 family genes to the module by examining the correlation of individual AP-1 family genes with module eigengene expression, finding that a number of the individual genes showed a strong correlation (Fig. 4.20A). I found high expression of these modules in all COVID-19 patient groups across the cell types, including recovery phase community cases, but lower expression in influenza and sepsis, suggesting specificity for COVID-19 (Fig. 4.20B). However, there was no consistent trend or correlation with COVID-19 severity or other clinical features, suggesting that the pathway was generic within COVID-19 illness.

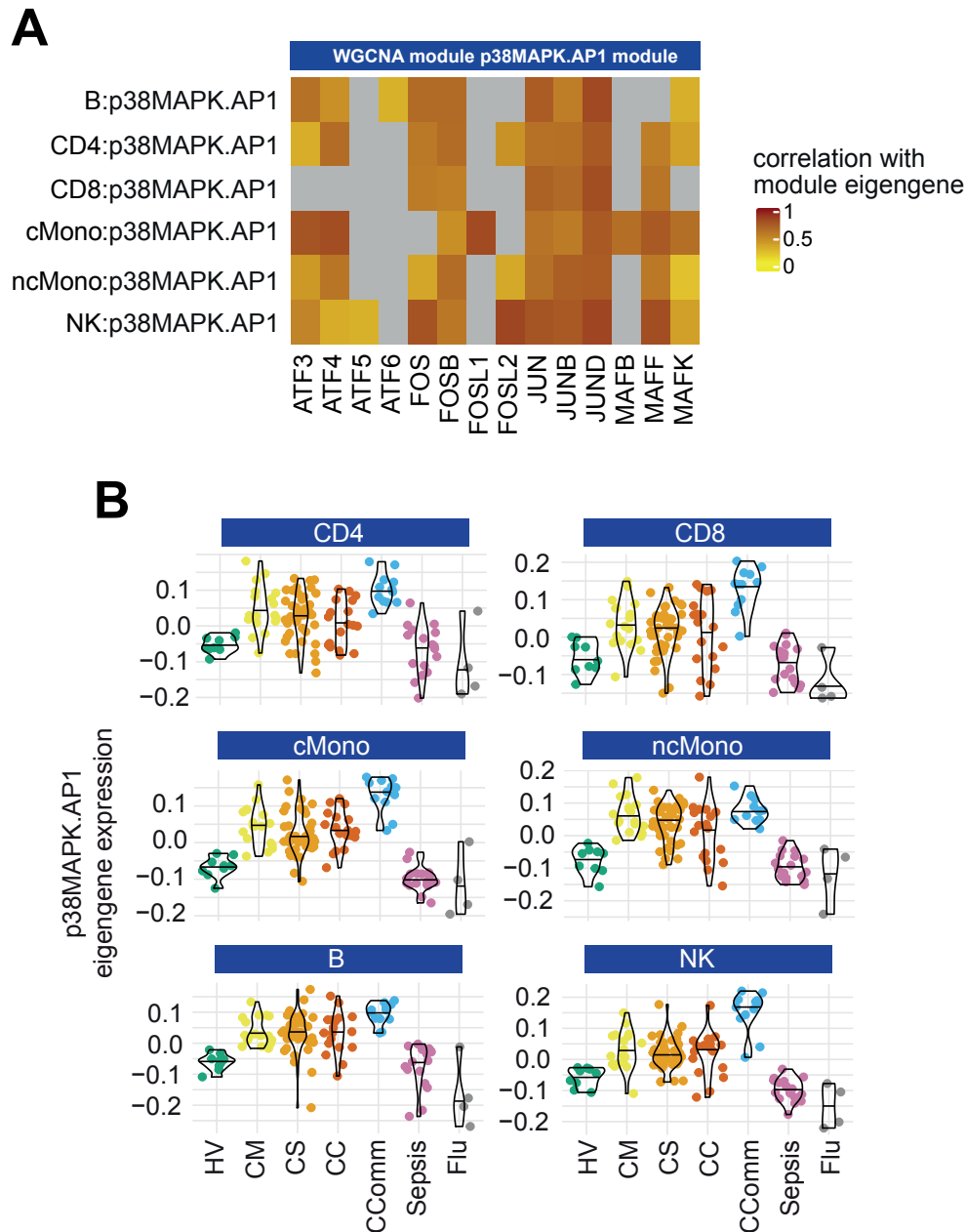
The third module set I found was distinguished by ribosome pathways and the KEGG 'COVID-19' pathway. Key genes within the module included genes relating to inflammasome function (including *NLRP1*, *MAP3K14* and *FOXP1*). Specifically, I also found that a relationship between COVID-19 severity and

expression of this module in cMonos, where there was negative correlation with disease severity (Fig. 4.21A).

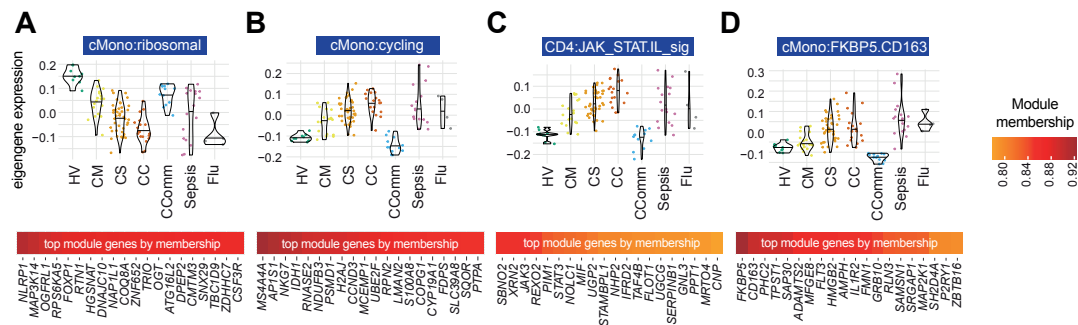
A fourth module set was enriched for classical (C2H2) zinc finger (ZNF) genes (Fig. 4.19A-B). This module was expressed more highly in HV/sepsis samples than in COVID-19/influenza, suggesting a degree of specificity to viral infection. This module set contained immune related genes such as *IRF2*, with a role in competitive inhibition of IRF1-mediated promotion of transcription of type I IFN genes, and *IL16*, important for chemoattracting CD4<sup>+</sup> expressing cells (Cruikshank et al. 2000).

The final common module set reflected cell cycling (“cycling” modules) based on pathway enrichments and module gene membership of *MKI67* and *TOP2A*, possibly reflecting the various cycling immune clusters identified from the detailed annotation analysis (Fig. 4.19A-B). There was a weaker correlation across cell type for these modules. The eigengene expression of the cycling cMono module increased with disease severity, in line with the increase in cycling cMono cluster proportion across disease severity seen in the differential abundance analysis (Fig. 4.16).

We also identified two cell-type-specific modules which were positively correlated to increasing COVID-19 severity, and were also elevated in sepsis and influenza. The first was a JAK-STAT/interleukin signaling module in CD4<sup>+</sup> T cells with pathway enrichment for IL-4, IL-6 and IL-13 signaling (Fig. 4.19A-B; Fig. 4.21C). The second was a module in cMono with EGFR pathway enrichment (Fig. 4.19A-B; Fig. 4.21D). This module contained genes including CD163 which modulates induced innate immune responses and acts as a scavenger receptor, as well as FKBP5 which influences glucocorticoid receptor sensitivity and is involved in the stress response.



**Figure 4.20: Weighted gene correlation network analysis (WGCNA) AP-1 related modules.** (A) Correlation of p38MAPK.AP1 module eigengene with genes of the AP-1 family. (B) Expression of p38MAPK.AP1 module eigengene across patient groups.



**Figure 4.21: Weighted gene correlation network analysis (WGCNA) modules with disease condition related expression patterns.** (A-D) Expression of module eigengene across patient groups for the (A) classical monocyte (cMono) ribosomal module (B) cMono cycling module (C) CD4+ T cell JAK-STAT interleukin signallign module and (D) cMono FKBP5.CD163 module. Top 20 genes per module by eigengene-gene correlations shown at the bottom of each module. Median eigenegene expression per module indicated by horizontal bar in violin plots.

### 4.3 Discussion

In this chapter, I presented a comprehensive PBMC cell census across multiple infection states ranging from COVID-19 of varying severity, severe influenza and all-cause sepsis. I showed multi-modal single cell sequencing data of an unprecedented scale, and the comprehensive multi-pronged multimodal approach successfully identified blood hallmarks of COVID-19. In particular, I showed the instances in which immune activation was specific to COVID-19 e.g. with AP-1 gene modules, vs. other occasions where they reflect a more generic immune response to severe infection e.g. cycling cMonos.

#### 4.3.1 Hallmarks of the immune response to COVID-19 and contrasts to other severe infections

##### Cell subset features

The results support the importance of MNP dysfunction in severe COVID-19 (Schulte-Schrepping et al. 2020). Hallmarks identified include proliferating cMono, high expression of antioxidant metallothionein and *S100A8/9* genes

which encode for calprotectin, together with reduced cDCs and pDCs. Since the time of data generation, other groups have reported similar findings (Schulte-Schrepping et al. 2020; Silvin et al. 2020). Upregulation of cMono and specifically higher abundance of S100A8/9/12 expressing cMono in severe/critical COVID-19 and sepsis is consistent with similar changes in monocytic myeloid-derived suppressor cells in both diseases, while similar reductions in ncMono and DCs were seen in both COVID-19 and sepsis. Together with the existing understanding of myeloid and MNP disturbances in sepsis (Monneret et al. 2006), this suggests a common myeloid pathway to the immune response to severe infection. There were also specific differences for other populations, including increased complement expressing (C1Q) ncMono in sepsis and extracellular matrix protein versican expressing cMono (mitohi/VCANhi) in COVID-19, which suggest that these may be features specific to COVID-19 beyond a general immune response to infection that would warrant further investigation for understanding COVID-19 specificities.

In terms of adaptive immunity, changes in the frequency of specific T cell subsets have been previously reported in severe COVID-19 (Sette and Crotty 2021; Rydzynski Moderbacher et al. 2020). Reports on T cell exhaustion in COVID-19 have been conflicting (Chen and John Wherry 2020; Kusnadi et al. 2021), as it has been difficult to dissect whether the expression of certain markers e.g. PD-1 represents T cell exhaustion or recent activation. In the analysis presented above, I found increases in proliferating effector T cells in both CD4+ and CD8+ compartments; however, similar to the limitations in other studies, it was not possible to distinguish whether the PD-1 expressed in these cell clusters reflected more of an activation or an exhaustion status. I chose to annotate these cells without the "exhaustion" label as I saw concomitant expression of other cell surface markers more suggestive of activation e.g. CD38 and HLA-DR, as well as genes suggesting that effector function was not lost e.g. *GZMB* and *GZMH*.

We saw increased numbers of plasmablasts across all disease states, in line with a common thread of adaptive immune responses across every infection condition. While there are differences between influenza and COVID-19 viral infection in terms of cell tropism and viral replication, there are also common themes of immune responses and mechanisms (Flerlage et al. 2021). The data presented here show similar changes in frequency of a variety of cell lineages in PBMCs COVID-19 and influenza including the majority of T and NK cell populations, HSCs, classical/cycling monocytes, and plasmablasts. However, I noted differences for both ncMonos and a number of B cell populations (differential naive, intermediate, and memory B cell response).

COVID-19 has been reported to predispose patients to thromboembolic disease, with reports of specific hyper-inflammatory thromboembolic phenotypes (Malas et al. 2020; Loo et al. 2021). This was also discussed in chapter 3 in terms of the pathways elevated in bulk RNA-seq in COVID-19 viral sepsis. The increase in platelet/CD34 megakaryocyte progenitors seen in this chapter may be related, considering the association in this dataset of the abundance of these cell types with thromboembolic occurrence. These findings are consistent with and substantiate data from other reports since the time of data generation of megakaryocytic lineage expansion in COVID-19 (Bernardes et al. 2020; Stephenson et al. 2021). To understand this phenomenon more fully, future experiments would likely have to include interrogation of the bone marrow to understand the lineage differentiation changes that are occurring in COVID-19 sepsis.

### **Gene expression changes**

In the gene expression space, I demonstrated substantial commonality of pathways across the disease conditions. This included gene modules involving JAK-STAT signaling and zinc finger proteins in both flu/COVID-19.



However, one of the most striking differences across disease conditions was the specific enrichment of AP-1/MAPK in COVID-19. One possibility is that COVID-19-induced downregulation of ACE2 may be involved in modulating both the renin-angiotensin system and activation of AP-1/p38MAPK signaling (Grimes and Grimes 2020), with specific evidence for spike protein induced AP-1 signaling (Patra et al. 2020). However, the lack of ACE2 expression in PBMCs (Muus et al. 2021) would suggest against the hypothesis of a spike protein related phenomenon, leaving the mechanisms behind AP-1 induction within PBMCs an open question.

An alternative hypothesis to the COVID-19 specificity of the AP-1 signaling, is that the observed module simply represents a feature of a naivety of an immune response, given that this was a novel pathogen. However, this hypothesis is difficult to interrogate given that it is difficult to find demographically and clinically appropriately matched control patient groups undergoing an immune response to a similarly novel pathogen. Given projections of more pandemics upcoming in this century, it is possible that similar studies in future outbreaks could be retrospectively compared to the current dataset to test this hypothesis.

#### **4.3.2 Opportunities from the CITE-seq atlas**

One of the initial aims from the experiments of this chapter was to create a publicly accessible data resource. This presents many unique opportunities for future re-use and re-analysis of the data.

First, as more single cell -omic datasets are generated across different disease conditions as tissue types, the need for cell type/state annotation for every dataset is both onerous and subject to investigator bias, making comparisons across studies difficult. Computational methods have recently advanced to leverage existing annotations and to reference map new (query) datasets onto existing (reference) datasets, including variational autoencoder and neural

network based methods, as well as bridge integration methods (Hao et al. 2021; Xu et al. 2021; Hao et al. 2022). These methods could be applied for future PBMC single cell sequencing datasets to map onto the detailed annotations of the current CITE-seq dataset for automated and aligned annotations.

Second, there are opportunities for integrative analysis of multi-modal data. There is currently a challenge in understanding how gene expression relates to protein levels, with a well known poor correlation between the two layers of biological information. Given the extensive panel of antibodies used in the CITE-seq dataset, the dataset can therefore be leveraged to build models of predicting protein expression from gene expression. Further opportunities for data integration are discussed in chapter 6.

Third, while the analysis presented has heavily focused on identifying active cell types/changes across disease states or understanding the pseudobulk changes per cell lineage, an interesting question which the CITE-seq dataset could help answering is what the changes of bystander adaptive immune cells are in the contexts of gross physiological alterations, specifically in this case severe infection. While naive lymphocytes (CD4/8 T cells and B cells) require receptor stimulation by specific antigen to undergo clonal selection and expansion, it is not well understood what the contributions to an immune response are of the lymphocytes which do not undergo such receptor ligation, but are nevertheless influenced by the surrounding cytokine milieu.

### **4.3.3 Limitations**

The approach and data described in this chapter have a number of limitations notwithstanding their comprehensive nature.

One analytical limitation to the results presented above is the lack of any trajectory analysis. The lack of spliced and unspliced counts from the preprocessing steps used precluded RNA velocity analysis, while the lack

of apparent immature cell types across the MNC compartments meant that forcibly fitting trajectories e.g. by principal curve methods risked producing artificial trajectories. In addition, the experimental design was such that maximizing unique patient numbers was prioritized, meaning that the cohort was underpowered for any time course trajectory analysis based on repeated sampling over time. A separate experiment to interrogate trajectories would be needed with a sample set designed to capture more time points over the course of patients' illness.

Another major limitation of the single cell profiling techniques presented above is the absence of functional wet-lab experimental validation, precluding conclusive insights into causal mechanisms. While the AP-1 pathway was identified as a distinguishing feature of the immune response to COVID-19, this could be further investigated by a combination of chromatin immunoprecipitation assays to understand the specific locations of AP-1 binding and factors within the AP-1 family that are relevant, as well as in vitro cell stimulation assays to characterize the stimuli that might be responsible for AP-1 activation in the different contexts. AP-1 has been shown to be linked to induction of type I IFN gene programs (Honda et al. 2006), but much remains to be understood regarding whether AP-1 may be specifically triggered by type I IFNs, and even if so, there would be a clear question of why the AP-1 signature in the current dataset is seen in COVID-19 patients but not in flu samples. Stimulation assays with pseudoviruses (COVID-19 and flu) may therefore assist in resolving this question.

## 4.4 Conclusions

Our understanding of immune responses to severe infection have improved substantially in recent years by interrogating peripheral blood genome wide gene expression. However, this has lacked the resolution to understand cell type

specific signatures of disease. The datasets and analysis presented in this chapter combine multiple modalities at a single cell resolution to compare the MNC immune responses across multiple severe infections, demonstrating remodeling in both innate and adaptive immune compartments and revealed cell specific shared and distinct aspects of the immune response to COVID-19 viral sepsis vs. other causes of sepsis.

# 5

## WHOLE BLOOD SINGLE-CELL MULTI-OMICS FOR SEPSIS SUBPHENOTYPES

---

*This chapter explores immune cellular basis underlying sepsis response signature transcriptomic subphenotypes.*

|     |                        |     |
|-----|------------------------|-----|
| 5.1 | Introduction . . . . . | 153 |
| 5.2 | Results . . . . .      | 161 |
| 5.3 | Discussion . . . . .   | 198 |
| 5.4 | Conclusions . . . . .  | 205 |

### 5.1 Introduction

As described in chapter 3, the heterogeneity in sepsis necessitates better patient stratification for understanding of the disease processes involved to allow targeted immunotherapies to be deployed. This chapter focuses on elucidating the drivers behind the various patient subgroups identified by the sepsis response signature (SRS) framework which was applied in chapter 3.

#### 5.1.1 Underlying bases of sepsis transcriptomic subphenotypes

Functional interpretation of the whole blood leukocyte transcriptome derived sepsis endotypes has thus far relied mostly on pathway analyses (via over-representation tests) of genes differentially expressed across endotypes.

In the original SRS description, SRS1 (the endotype associated with higher early mortality) was enriched in signatures defining endotoxin tolerance and T cell exhaustion, as well as showing a downregulation of pathways related to antigen presentation (especially HLA class II) and T cell activation (Davenport et al. 2016). Network analysis of the genes with increased expression in SRS1 also revealed networks with transcription factors HIF1 $\alpha$  and HIF2 $\alpha$  as nodes, notable given previous observations of the importance of HIF1 $\alpha$  in reprogramming monocytes in sepsis (Shalova et al. 2015). However, as the original patients from which SRS was derived consisted only of community acquired pneumonia (CAP) patients, the relevance of the HIF proteins is somewhat confounded by the fact the patients who were more ill would likely have more impaired respiratory function which may or may not have been detectable clinically. Indeed, pathway analysis in fecal peritonitis patients (FP) where SRS was defined did not reveal the same HIF related networks (Burnham et al. 2017).

In the MARS classification, the Mars1 endotype which was the one consistently associated with the highest mortality also displayed reduced expression of genes related to immune function (both innate and adaptive arms) including antigen presentation, T-cell receptor pathways as well as Toll-like receptor (TLR) and NF- $\kappa$ B signaling (Scicluna et al. 2017). Sweeney et al (2018) uncovered three endotypes termed "Inflammopathic" (higher mortality), "Adaptive" (lower mortality) and "Coagulopathic" (also higher mortality). These endotypes showed enrichment in innate immunity (Complement activity, and IL-1 receptor and pattern recognition receptor signaling), adaptive immunity and clotting/coagulation (platelet degranulation and the coagulation cascade) pathways respectively.

More recently, Baghela et al (2022) stratified sepsis patients into "Neutrophilic-Suppressive", "Inflammatory", "Innate-Host-Defense", "Interferon" and "Adaptive" groups, with the first two having the highest 28 day mortality.

The authors showed overlap of these two signatures with the SRS1 and Mars1 groups from previous definitions, suggesting that similar axes of variation were identified. In both of the high mortality groups in this study, there was enrichment of T and B cell signaling pathways in genes downregulated in the groups.

Common themes arising from these studies are therefore the reduction in adaptive immunity pathways and antigen presentation in transcriptomic subphenotypes of worse outcome. However, the mechanisms relating to pathways that are elevated in poor outcome groups remain mostly unresolved beyond understanding that it involves innate immune signaling pathways and cells. Moreover, it is difficult from bulk transcriptomic data to understand whether observed differences are due to cell type compositional changes, or differential gene expression within specific cell types. Analyses of the SRS subphenotypes in the original report (Davenport et al. 2016) did indeed take clinical cell counts into consideration, as did analyses presented in chapter 3, suggesting that there are genuine gene expression differences beyond changes in cellular abundance.

### **5.1.2 Whole blood single cell multi-omics**

While the data presented in chapter 4 moved beyond bulk transcriptomics to achieve single cell resolution, the cell types examined were restricted to the peripheral blood mononuclear cells (PBMCs), with the exclusion of granulocytes which form the majority of peripheral blood leukocytes. To understand the cellular basis of SRS transcriptomic subphenotypes, the most direct way would be to perform scRNA-seq of whole blood samples from sepsis patients given that the signatures were derived from whole blood samples. However, this brings the challenge of handling neutrophils in the laboratory, as these fragile cells are not amenable to cryopreservation or fixation, necessitating immediate processing

every time blood is drawn from a patient (Agashe et al. 2017; Braudeau et al. 2021; Hope et al. 2021).

While the scRNA-seq methodology has begun to be employed in interrogating sepsis in general (Reyes et al. 2020; Wen et al. 2020), these studies have likewise focused on peripheral blood mononuclear cells (PBMCs) and did not examine the contribution of neutrophils. Moreover, the studies have mostly considered differences between sepsis and healthy or other inflammatory states e.g. non-sepsis critical illness (Reyes et al. 2020), but did not attempt to resolve heterogeneity within sepsis.

Processing fresh samples presents both logistical challenges, and an issue of potential confounded batch effects where each sample represents a single batch. However, given the lack of alternatives to store neutrophil cellular and nucleic acid material, fresh profiling for scRNA-seq was the only solution, motivating the experiments in this chapter of fresh sample whole blood single cell multi-omics.

As in the previous chapter, I sought to obtain multi-modal data from the single cell experiments. The extension of commercial droplet based platforms i.e. from 10X of scRNA-seq to neutrophils has occurred recently (Wigerblad et al. 2022; Qi et al. 2021; Sinha et al. 2021), but an alternative technology employed by the BD Rhapsody system that allows gentle capture of cells based on a gravity sedimentation-microwell approach (Gao et al. 2020) has also been deployed to effectively capture neutrophils (Schulte-Schrepping et al. 2020). In a head to head comparison, it was shown that the Rhapsody system had superior performance to the 10X system in capturing granulocyte transcripts (Qi et al. 2021). Moreover, the Rhapsody system also permits multi-modal profiling in terms of simultaneous cell surface and RNA profiling and as such was used as the method of choice for the experiments in this chapter.



### 5.1.3 Neutrophils

Neutrophils are the human body's first line of defence against invading pathogens. The most abundant circulating leukocyte, neutrophils are commonly accepted to have a short half life on the order of hours-days (Summers et al. 2010) although human in vivo deuterium labeling studies have demonstrated a half life closer to one week (Pillay et al. 2010).

Morphologically, neutrophils are defined by lobed nuclei and the presence of various granules containing proteases and other molecules with potent antimicrobial properties. Upon injury, infectious or sterile, neutrophils are attracted to the site most notably by lipids such as leukotriene B<sub>4</sub>, complement anaphylatoxins (especially C<sub>5a</sub>), chemokines (chiefly CXCL<sub>8</sub>) and formyl peptides as N-formylmethionine-leucyl-phenylalanine (Metzemaekers et al. 2020). Given that severe infections can cause consumption of these cells in large numbers, there is a large reserve of neutrophils in the bone marrow for rapid replenishment of circulating cells (Donohue et al. 1958); in mice, it is estimated that there are 120 million neutrophils in the bone marrow in contrast to the <2.5 million in circulation (Furze and Rankin 2008).

In humans, it is estimated that  $\sim 10^{11}$  neutrophils are released from the bone marrow every day (Mauer et al. 1960). Neutrophil mobilization from the bone marrow depends on a balancing interaction between CXCR<sub>2</sub> and CXCR<sub>4</sub> signaling (Summers et al. 2010). While osteoblasts and stromal cells within the marrow produce CXCL<sub>12</sub> which binds to CXCR<sub>4</sub> for neutrophil retention, G-CSF interferes with this axis by reducing expression of both CXCL<sub>12</sub> (Semerad et al. 2002; Semerad et al. 2005) and CXCR<sub>4</sub> (Hyun et al. 2006; Devi et al. 2013).

Once mobilized from the bone marrow and recruited to the site of pathology, neutrophils deploy their effector mechanisms to combat invading pathogens. This includes degranulation, phagocytosis and neutrophil extracellular trap

(NET) formation (Burn et al. 2021). Phagocytosis involves physically engulfing the pathogen into a phagosome followed by phagosome maturation during which granules containing antimicrobial molecules fuse with the phagosome for microbial killing (Nordenfelt and Tapper 2011). There is also an NADPH oxidase complex in phagosome membranes which generates reactive oxygen species e.g. superoxide, which in turn can stimulate myeloperoxidase for hypochlorite production (Valenta et al. 2020). Degranulation involves releasing antimicrobial soluble proteins into the extracellular space instead of fusion with phagocytosed microbes. NETosis is a process where chromatin decondensation occurs, and granule effector proteins are attached followed by extrusion (Burgener and Schroder 2020) to adhere to, capture and kill pathogens.

#### **5.1.4 MDSCs**

Myeloid derived suppressor cells (MDSCs) is a term introduced just over a decade ago (Gabrilovich et al. 2007), originally used in the context of tumor biology to describe immune cells of myeloid lineage with immunosuppressive function. The controversy arising from this stemmed mainly from the lack of either specific morphological or phenotypic markers that could unequivocally identify these cells as a specific cell type (Veglia et al. 2018), necessitating functional readouts of suppressive function. This has usually been assessed by T cell suppression in various forms of co-culture assays, either in an antigen specific or non-specific format (Bronte et al. 2016). This itself is fraught with difficulty in interpretation of MDSC relevance in vivo given the difficulty in measuring suppression in vivo. Furthermore, as phenotypic readouts are frequently used to claim identification of MDSCs, it is difficult to assess whether many reported MDSCs would truly be suppressive.

MDSCs are thought to arise from persistent myeloid compartment stimulation, especially in the form of myelopoeitic stimuli and other inflammatory signals

e.g. GMCSF, MCSF, IL-6, and IL-1 $\beta$  (Veglia et al. 2021). MDSCs can be split into either granulocytic or monocytic (G-MDSCs / M-MDSCs). In general, both types display features of immature cell surface phenotypes (Veglia et al. 2018); however, G-MDSCs with a mature phenotype have been described in the setting of EL-4 tumor mouse models (Youn et al. 2012) and in Hodgkin and non-Hodgkin lymphoma in humans (Marini et al. 2016). In the laboratory, G-MDSCs have often been isolated by obtaining the low density neutrophil fraction i.e. neutrophils which are obtained from the PBMC fraction after density gradient centrifugation. Recent data from human non-small cell lung cancer and head and neck cancer samples identified LOX-1 as a marker that enriches for G-MDSCs more than via a density definition (Condamine et al. 2016). Transcriptionally, MDSCs have also begun to be defined e.g. by Kessenbrock and colleagues who sequenced 14,646 cells in a mouse model of breast cancer to find 105 genes which were commonly expressed and specifically elevated in both G and M-MDSCs (Alshetaiwi et al. 2020). These gene sets were then shown to be enriched in human breast cancer neutrophil samples as well. Crucially in this study, the authors validated the suppressive nature of both G and M-MDSCs by co-culturing these cells with T cells, demonstrating suppression of both CD4+ and CD8+ T cell proliferation.

### **5.1.5 Neutrophils in sepsis and other infectious diseases**

Neutrophils have previously been studied in the setting of sepsis. In cecal ligation and puncture (CLP) mouse models, a mouse model of sepsis, neutrophils were shown to have impaired chemotaxis and PD-L1 mediated suppression of T cells (Qi et al. 2021) and impaired respiratory burst function as well reduced levels of cell surface CD11b expression (Grailer et al. 2014), a marker of neutrophil maturity. Other studies have identified a general expansion of MDSCs in sepsis (Delano et al. 2007), as well as their mobilization through liver acute phase proteins (Sander et al. 2010).

Indeed, the notion of immature neutrophils in sepsis has long been known, understood clinically as a "left shift" upon examining peripheral blood films and cell counts in septic patients whereby immature neutrophils (e.g. band cells) are increased. Human immunophenotyping studies are consistent with the pre-clinical data and clinical observations, with immature neutrophils associated with mortality in sepsis (Daix et al. 2018) and reduced CD88 (marker of C5a mediated neutrophil dysfunction (Conway Morris et al. 2009; Morris et al. 2011) associated with risk of developing secondary infections (Conway Morris et al. 2018). In the ExPRES-sepsis cohort study, Walsh and colleagues found using flow cytometry that increased neutrophil cell surface CD24 and CD279 were among the best biomarkers for predicting deterioration of patients with suspected infection to sepsis (Shankar-Hari et al. 2018), with CD24 relevant for delaying neutrophil apoptosis and CD279 relevant for its potential role in mediating T cell exhaustion (Chang et al. 2014).

Given the relevance of immature neutrophils in sepsis, it is natural to consider whether these cells may behave as MDSCs. This has been considered in the context of risk of nosocomial infection amongst sepsis patients, where it was shown that compared to healthy donors, both M-MDSCs (CD14<sup>+</sup> HLA-DR<sup>lo</sup>/negative) and G-MDSCs (CD14<sup>-</sup>CD15<sup>+</sup>) were expanded in septic and non-septic critically ill patients, but G-MDSCs were specifically higher in the sepsis patients (Uhel et al. 2017). Sepsis patients with a higher fraction of G-MDSCs had a higher rate of nosocomial infections (Uhel et al. 2017).

### **5.1.6 Aims**

The overall aim of this chapter is to dissect the cellular immunological basis of previously reported peripheral blood bulk transcriptomic endotypes of sepsis. Specifically, I aim to:

1. generate a whole blood multi-omic single-cell atlas for sepsis, comparing

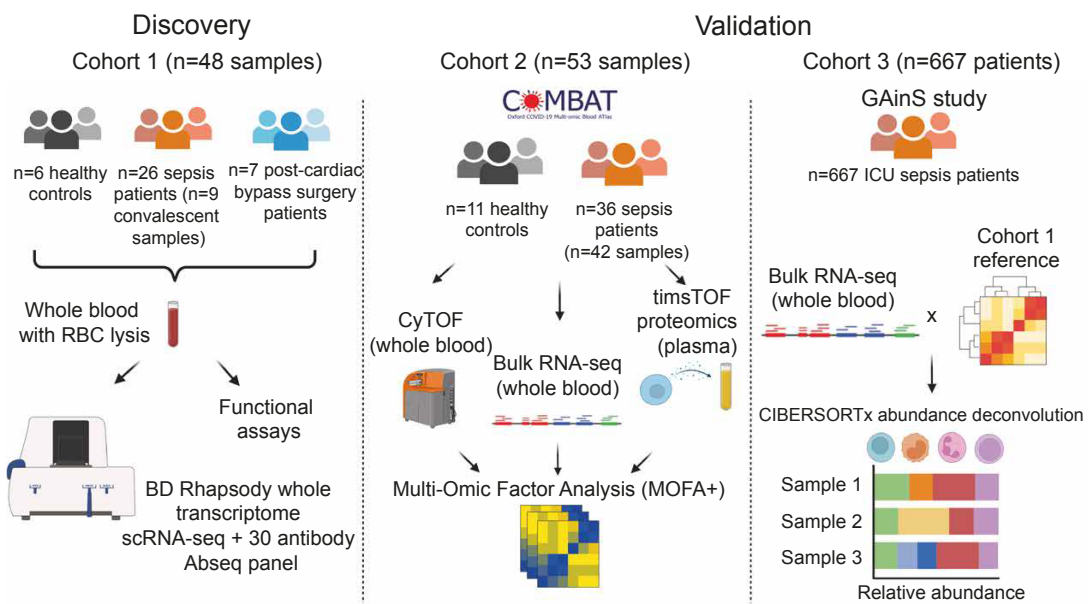
with healthy volunteers and the sterile immune response (patients undergoing elective cardiac surgery)

2. investigate the role of neutrophils in sepsis including specific neutrophil subpopulations and immunosuppressive function
3. determine cellular drivers, mediators and pathways of SRS and the role of neutrophils in this from the whole blood multi-omic single-cell atlas
4. validate the identified drivers of SRS in integrative analysis of mass cytometry (CyTOF) and bulk RNA-seq data from separate sepsis cohorts

## **5.2 Results**

### **5.2.1 Clinical cohort**

In this chapter, I describe data and analysis from three cohorts. Cohort 1 consisted of 26 all-cause sepsis patients, 6 healthy volunteer (HV) controls and 7 post cardiac surgery (CS) samples as sterile inflammation controls for comparison to the sepsis patients. Sepsis and post cardiac surgery patients were recruited under the Sepsis Immunomics study. Cohort 2 included the sepsis and HV samples from chapters 3 and 4. Cohort 3 included 864 adult ( $\geq 18$  years old) sepsis samples ( $n=667$  individuals). The profiling strategy is depicted in Figure 5.1 with clinical details given in tables 5.1, 5.2 and 5.3.



**Figure 5.1: Approach for understanding cellular basis of whole blood SRS signatures.** Patient cohort, experimental and analytical strategies for discovery cohort (cohort 1, whole blood single cell RNA and cell surface protein profiling), COMBAT sepsis samples from chapters 3 and 4 (cohort 2, whole blood bulk RNA-seq and CyTOF and plasma proteomic data) and GAinS study sepsis samples (cohort 3, whole blood bulk RNA-seq).

| Trait                                       | Sepsis cohort (n = 26) | Cardiac surgery cohort (n = 7) | Healthy volunteer cohort (n = 6) | FDR  |
|---------------------------------------------|------------------------|--------------------------------|----------------------------------|------|
| Age                                         | 63 (17)                | 60 (16)                        | 63 (22)                          | 0.93 |
| Male sex                                    | 16 (62%)               | 5 (71%)                        | 3 (50%)                          | 0.93 |
| SOFA score                                  | 4.8 (3.6)              | -                              | -                                | -    |
| Intensive care (at point of sampling)       | 14 (54%)               | -                              | -                                | -    |
| Mechanical ventilation                      | 5 (19%)                | -                              | -                                | -    |
| Vasopressors                                | 8 (31%)                | -                              | -                                | -    |
| Renal replacement therapy                   | 1 (4%)                 | -                              | -                                | -    |
| White cell count                            | 15.5 (6.4)             | -                              | -                                | -    |
| Proportion neutrophils                      | 82%                    | -                              | -                                | -    |
| Proportion lymphocytes                      | 9%                     | -                              | -                                | -    |
| Proportion monocytes                        | 6%                     | -                              | -                                | -    |
| Proportion eosinophils                      | 1%                     | -                              | -                                | -    |
| Days from hospital admission until sampling | 2 (1.5)                | -                              | -                                | -    |

**Table 5.1: Clinical characteristics of cohort 1 (n = 39 individuals).** Kruskal-Wallis tests performed for continuous data and chi-square test performed for count data. Data are n (%) or mean (SD) unless otherwise specified.

| Microbiology                                    | n (%)    |
|-------------------------------------------------|----------|
| <i>E. coli</i>                                  | 4 (15%)  |
| <i>E. coli</i> + <i>Clostridium perfringens</i> | 1 (4%)   |
| <i>S. pneumoniae</i>                            | 1 (4%)   |
| <i>S. aureus</i>                                | 2 (8%)   |
| <i>Proteus mirabilis</i>                        | 2 (8%)   |
| Coagulase negative staphylococcus               | 1 (4%)   |
| <i>Pseudomonas aeruginosa</i>                   | 1 (4%)   |
| <i>Citrobacter</i>                              | 1 (4%)   |
| Unknown                                         | 13 (50%) |

**Table 5.2: Microbiology characteristics of cohort 1 sepsis patients (n = 26 individuals).**

### 5.2.2 Generating a fresh whole blood multi-omic single cell atlas

I first profiled the whole transcriptome and 30 cell surface proteins (chapter 2) of single cells from fresh, red cell lysed, whole blood samples without other perturbation to the white blood cell compartment, ensuring an unbiased sample type context for the single cell atlas. After quality control (chapter 2), I retained 272,993 cells and 23,103 genes for downstream dimensionality reduction (DR), clustering, visualisation, annotation and differential analysis.

I performed initial broad annotation of the dataset by first manually annotating clusters based on lineage specific protein marker expression (Fig. 5.2), followed by cross referencing annotations with a data driven, algorithmic approach where annotations per cell were assigned based on reference pure immune cell type RNA-seq profiles from the Blueprint Consortium (Fig. 5.3). I found the two approaches to be broadly concordant.

At this level of annotation, I recapitulated well described hallmarks of sepsis peripheral blood immunophenotypes including neutrophilia, lymphopenia and reduced classical monocyte (cMono) HLA-DR gene and surface protein expression (Fig. 5.4). However, the same phenotypes are seen in CS samples, suggesting that the changes reflect a general inflammatory insult rather than



| Infection source             | n (%)   |
|------------------------------|---------|
| Community acquired pneumonia | 9 (35%) |
| Urosepsis                    | 5 (19%) |
| Intra-abdominal sepsis       | 8 (31%) |
| Meningitis                   | 1 (4%)  |
| Necrotising fasciitis        | 1 (4%)  |
| Infective endocarditis       | 1 (4%)  |
| Bacteraemia                  | 1 (4%)  |

Table 5.3: Septic source characteristics of cohort 1 sepsis patients (n = 26 individuals).

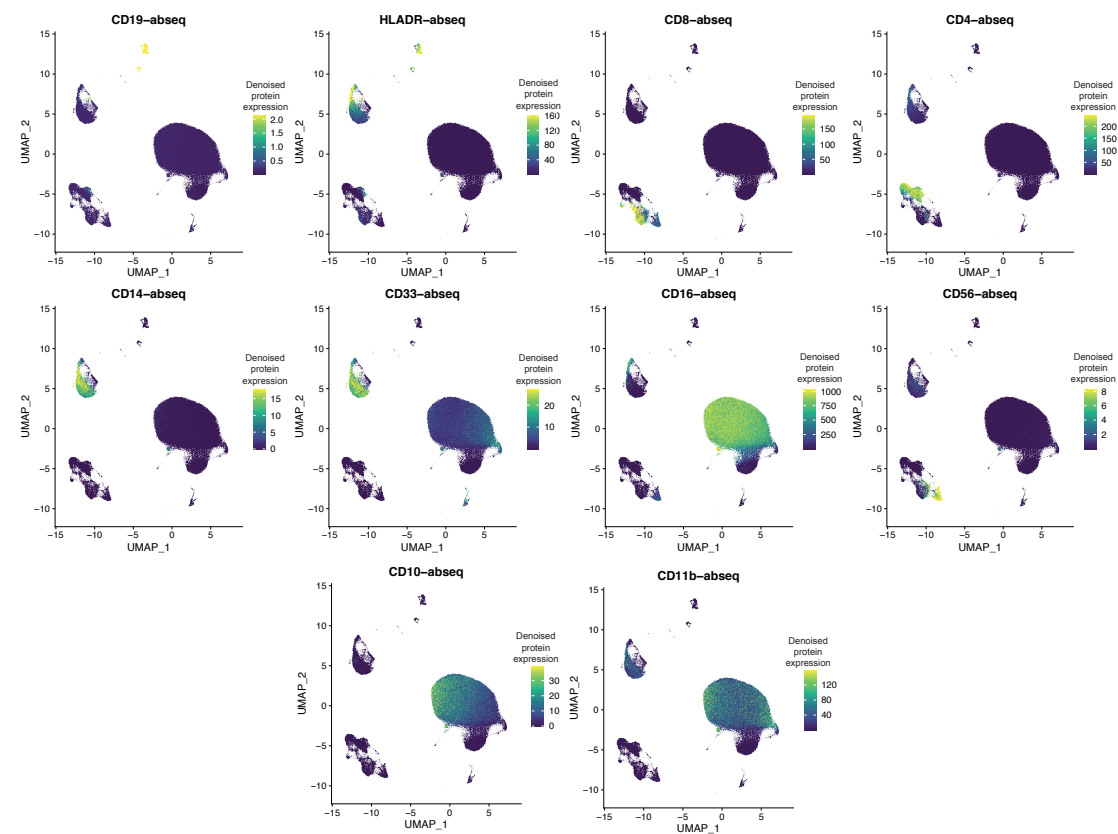
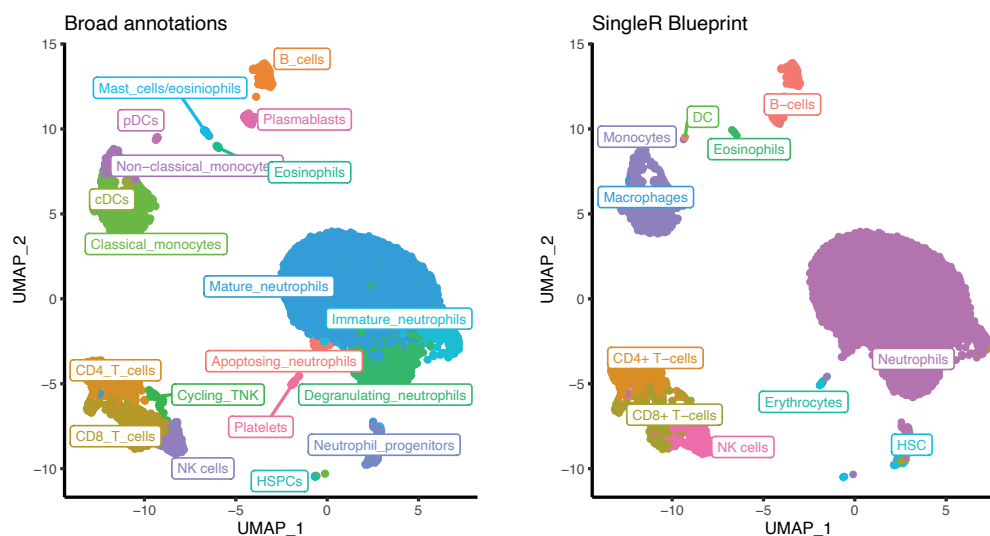


Figure 5.2: Cell surface lineage markers cohort 1 single cell RNA and cell surface protein sequencing dataset. UMAP of 272,993 cells with normalized surface protein marker expression overlaid for selected lineage defining markers.



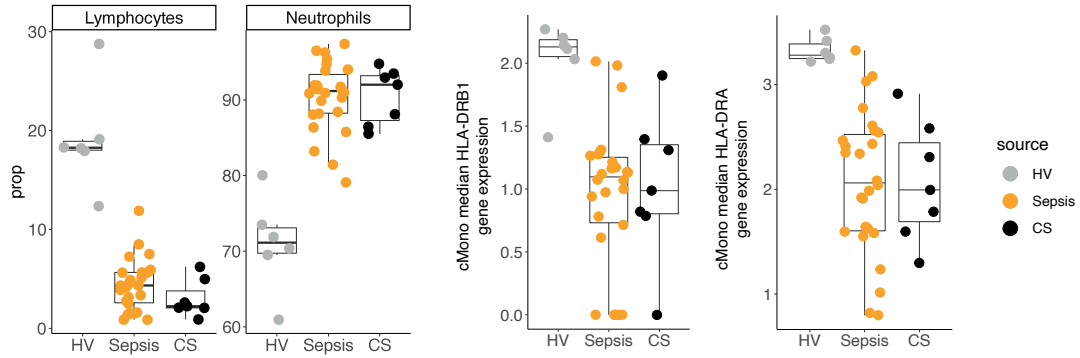
**Figure 5.3: Broad annotation of cohort 1 single cell dataset.** UMAP of 272,993 cells comparing broad manual annotations (left, chapter 2 “Whole blood single cell multi-omic analysis (cohort 1)”) with data driven algorithmic labeling of cell state/identity by SingleR based on reference bulk RNA-seq profiles of pure cell populations from the Blueprint Consortium (right).

being specific to sepsis.

### 5.2.3 Identifying immature neutrophils in the atlas

Upon clustering for finer grained annotation (Fig. 5.5A; Appendix), I noted the presence of multiple neutrophil clusters (Fig. 5.5B). CD10 expression visualized on uniform manifold approximation and projection (UMAP) suggested the presence of neutrophils at varying stages of maturation (Fig. 5.2), in line with previous studies of neutrophil single cell analysis of other disease states.

I examined the gene markers for these neutrophil clusters and observed multiple genes suggesting immaturity as previously reported (Sinha et al. 2021). This included *CD24* and *ARG1*, as well as genes encoding for neutrophil granule proteins such as *MPO*, *ELANE*, *AZU1*, *CTSG*, *RNASE2* (Fig. 5.6A-B). I annotated the clusters according to gene markers that particularly distinguished the neutrophil state, including a cluster of neutrophils with marked *IL1R2*



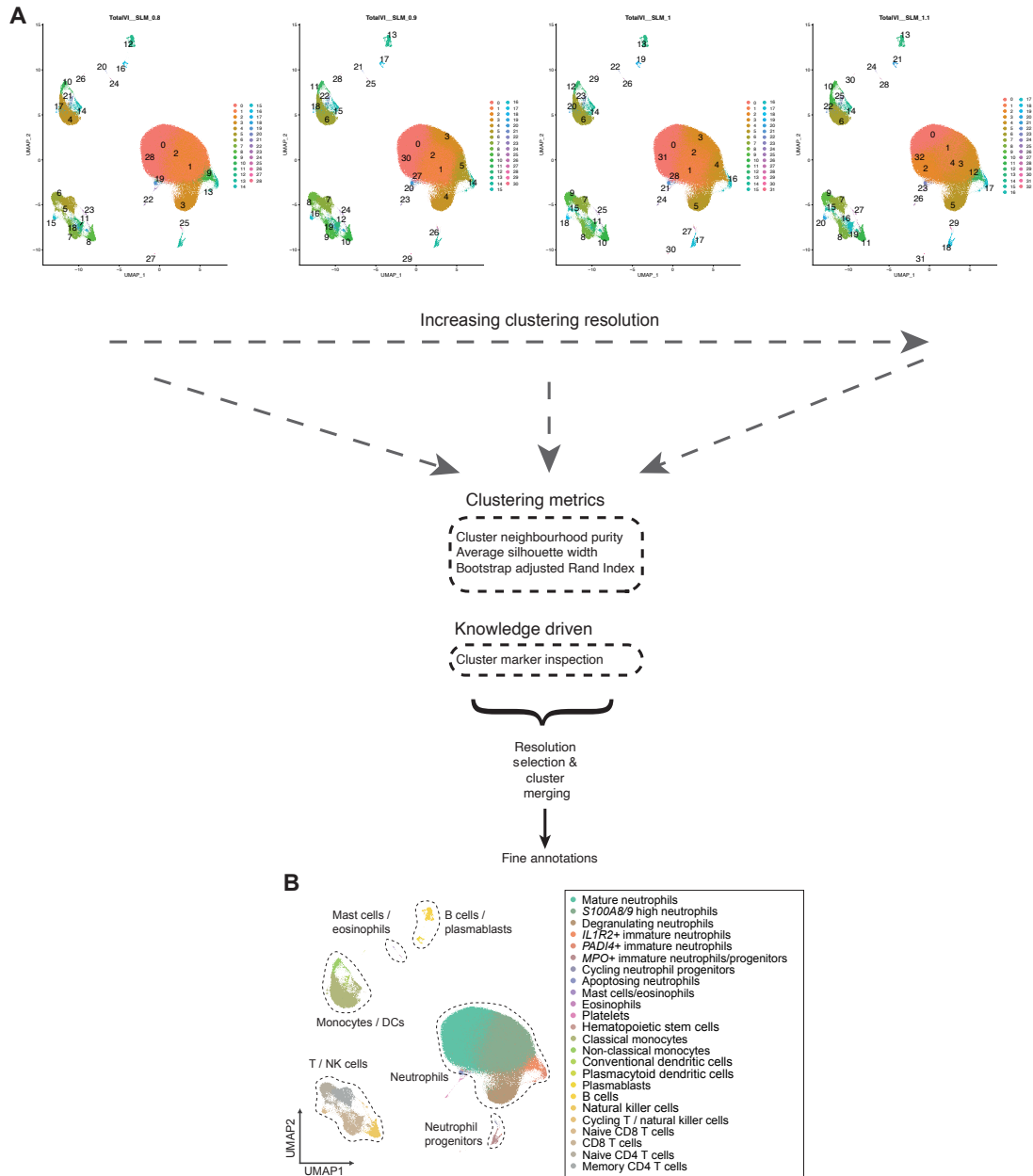
**Figure 5.4: Peripheral blood immune hallmarks of sepsis and systemic inflammation.** Boxplots of proportions of neutrophils and lymphocytes and HLA-DRA and HLA-DRB1 gene expression in classical monocytes in healthy volunteer (HV, n=6), sepsis (n=26) and post cardiac surgery (CS, n=7) samples from cohort 1.

expression and a population undergoing cell cycle (Fig. 5.6C-D).

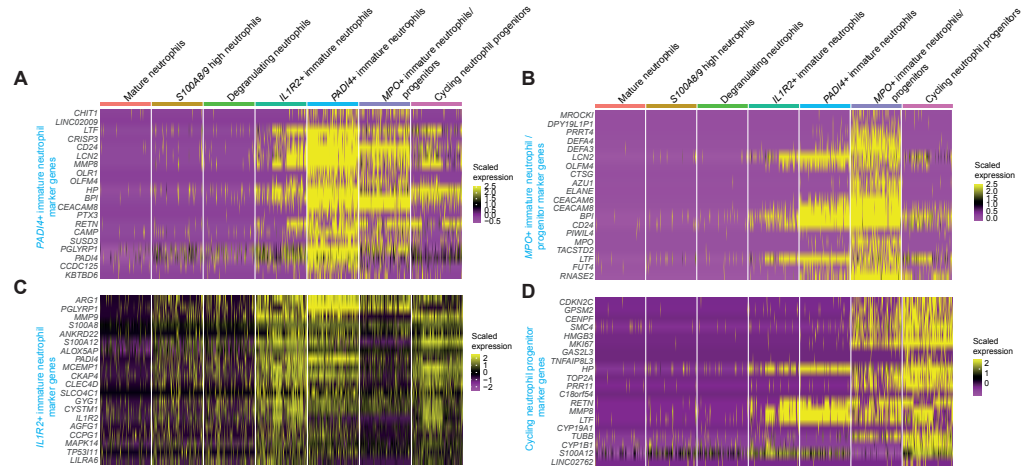
For analytical validation of the maturity annotations, I performed RNA velocity analysis on the neutrophil compartment for data driven identification of differentiation trajectories. I found a velocity direction starting from clusters annotated as progenitors (Fig. 5.7A), forming a single overall trajectory. The velocity length of the velocity vector field was substantially greater in the areas of the immature neutrophil clusters, suggesting that most of the differentiation stopped once either the S100A8/9 high neutrophil state or mature neutrophil state were reached (Fig. 5.7B). Partition based graph abstraction analysis (Wolf et al. 2019) supported a sequential progression of maturation according to the observed clusters, with two terminal states of differentiation into the annotated mature neutrophils and S100A8/9 high neutrophils (Fig. 5.7C).

#### 5.2.4 Delineating cell populations specific to sepsis

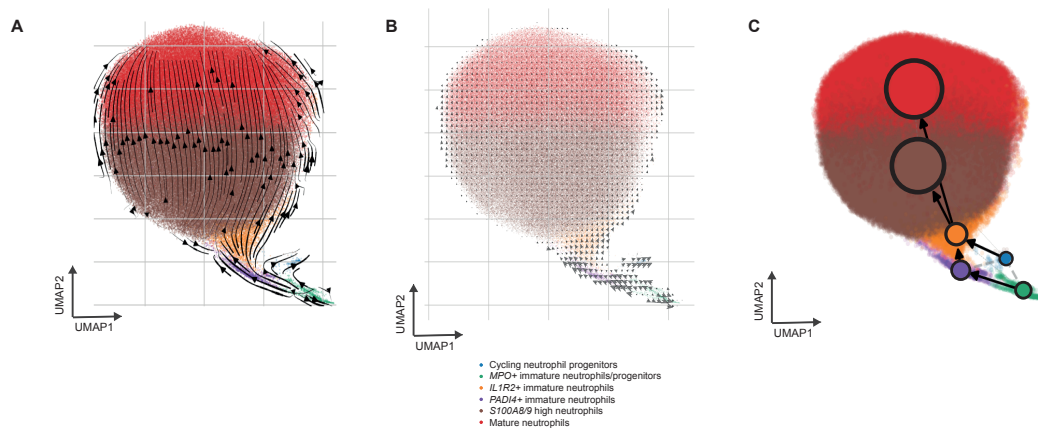
I next investigated the immune cell compositional changes in the septic state at the finer resolution of immune cell annotations. I utilized a graph neighborhood based method to sample neighborhoods in the single-cell dataset to model cell states/type changes across conditions. This method improves sensitivity



**Figure 5.5: Fine clustering approach for cohort 1 single cell dataset.** (A) Clustering results at varying resolutions with smart local moving algorithm for (B) fine annotation strategy (Appendix).

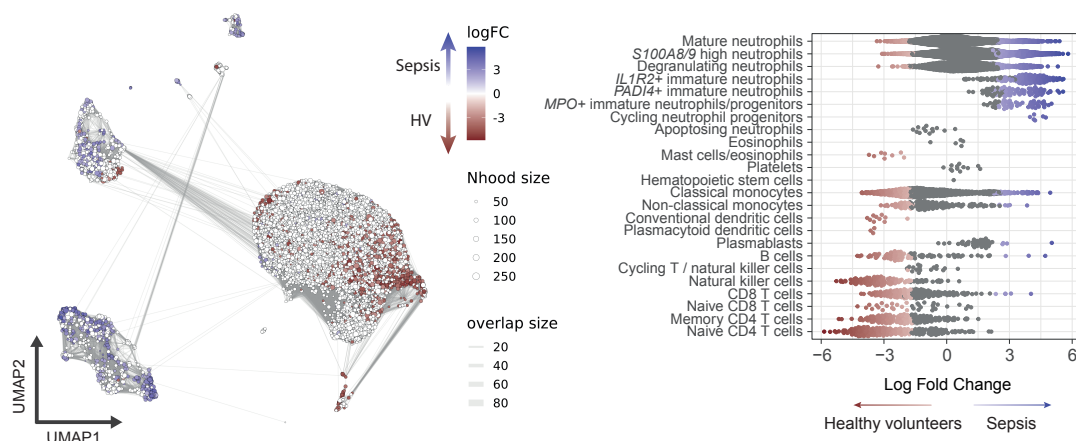


**Figure 5.6: Cohort 1 immature neutrophil subset markers.** Heatmaps of gene markers for (A) *PADI4*<sup>+</sup> immature neutrophils (B) *IL1R2*<sup>+</sup> immature neutrophils (C) *MPO*<sup>+</sup> immature neutrophils/progenitors and (D) cycling neutrophil progenitors from cohort 1 neutrophils.



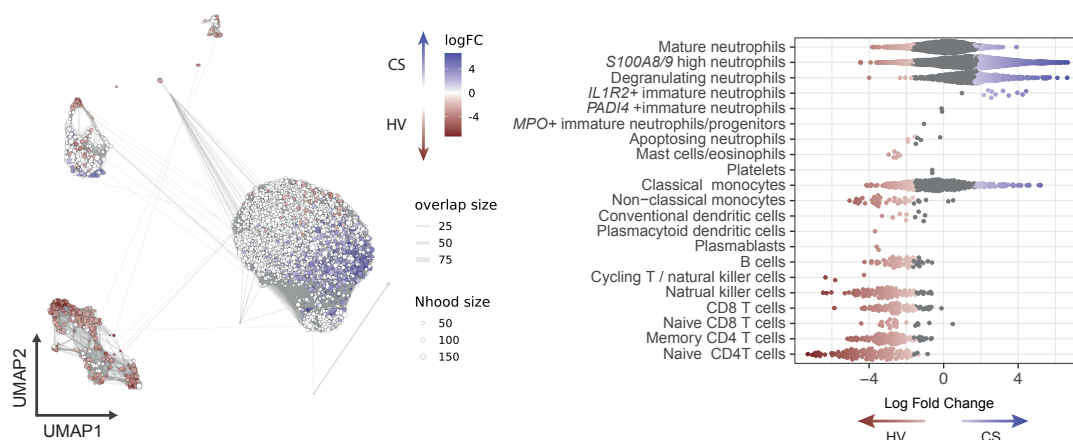
**Figure 5.7: Trajectory based analytical validation of cohort 1 neutrophil maturity status.** (A) UMAP of cohort 1 neutrophils (excluding degranulating neutrophils) with RNA velocity stream directions plotted. (B) UMAP as per (A) with size of velocity vector throughout vector field depicted. (C) Partition based graph abstraction analysis of RNA velocity predicted cellular trajectory transitions.

and mitigates loss of signal due to discretizing cells into clusters (Dann et al. 2021). I first compared differential abundance between sepsis and HV samples. Neighborhoods in multiple cell states differed significantly in proportion, with all immature neutrophil populations increasing in sepsis vs. HV, as well as degranulating and S100A8/9 high neutrophils (Fig. 5.8). By contrast, cells enriched in HVs included virtually all mononuclear cell (MNC) subsets except plasmablasts. A similar pattern of compositional remodeling could be observed when comparing CS patients to HVs (Fig. 5.9).

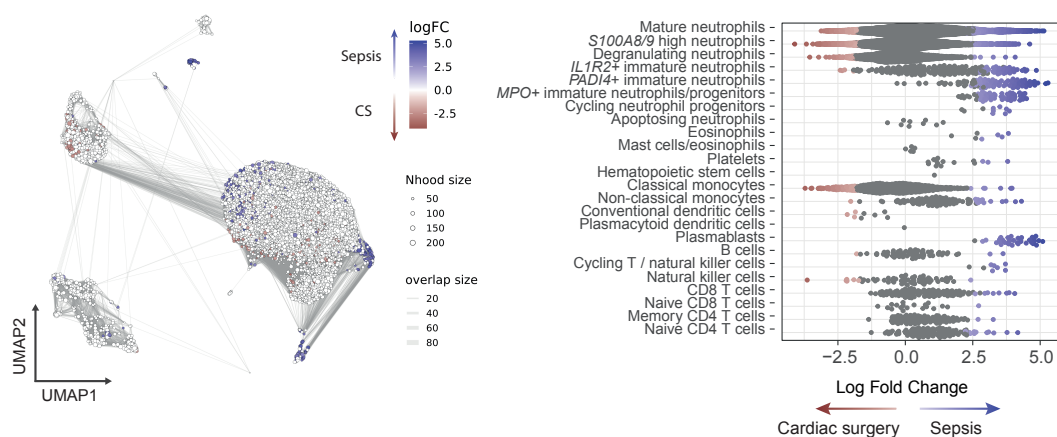


**Figure 5.8: Cohort 1 cellular composition analysis between sepsis and healthy volunteers (HV).** Differential abundance (DA) analysis using graph neighborhood (Nhoo) based method to detect enrichment of neighborhoods between sepsis (n=26) and HV (n=6) samples, with UMAP of sampled neighborhoods and statistically significant enrichment colored (spatial FDR <0.05 with generalized linear modeling). (B) Beeswarm plot of sepsis-HV DA analysis with cluster labels of neighborhoods depicted.

When contrasting sepsis patients with CS patients, I found significant enrichment of the four immature neutrophil populations in sepsis (Fig. 5.10). Thus, whereas the common hallmarks of sepsis in terms of neutrophilia, lymphopenia and low cMono HLA-DR may in fact be more reflective of a general inflammatory insult (Fig. 5.4), immature neutrophils appeared to be specific to the septic peripheral blood immune cell composition and not seen in the sterile inflammatory condition of post cardiac surgery.



**Figure 5.9: Cohort 1 cellular composition analysis between post cardiac surgery sterile inflammation (CS) and healthy volunteers (HV).** Differential abundance (DA) analysis using graph neighborhood (Nhood) based method to detect enrichment of neighborhoods between CS (n=7) and HV (n=6) samples, with UMAP of sampled neighborhoods and statistically significant enrichment colored (spatial FDR <0.05 with generalized linear modeling). (B) Beeswarm plot of CS-HV DA analysis with cluster labels of neighborhoods depicted.



**Figure 5.10: Cohort 1 cellular composition analysis between sepsis and post cardiac surgery sterile inflammation (CS).** Differential abundance (DA) analysis using graph neighborhood (Nhood) based method to detect enrichment of neighborhoods between sepsis (n=26) and CS (n=7) samples, with UMAP of sampled neighborhoods and statistically significant enrichment colored (spatial FDR <0.05 with generalized linear modeling). (B) Beeswarm plot of sepsis-CS DA analysis with cluster labels of neighborhoods depicted.



### 5.2.5 Immunosuppressive neutrophils in sepsis

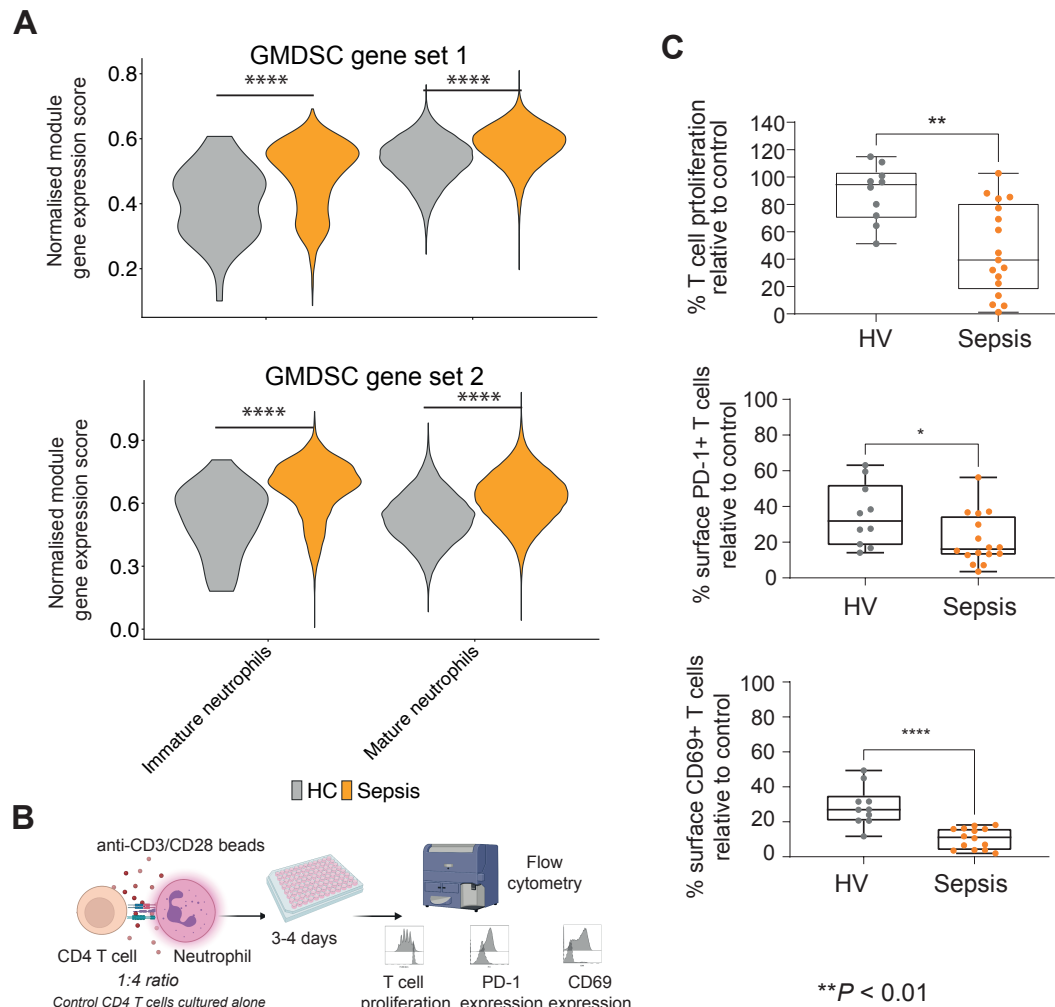
Immature neutrophils have frequently been suggested to be suppressive but many reports tie these two phenomena relying purely on phenotypic observation (Schulte-Schrepping et al. 2020). I scored all neutrophil populations for expression of published gene sets derived from experimentally validated immunosuppressive neutrophils (granulocytic myeloid derived suppressor cells, G-MDSCs) (Alshetaiwi et al. 2020; Bayik et al. 2020), finding a significantly higher enrichment score of these gene sets in both mature and immature neutrophils from sepsis compared to those from HVs (Wilcoxon rank sum tests, FDR <0.0001) (Fig. 5.11A).

To validate this experimentally, I designed an in vitro neutrophil – allogeneic CD4<sup>+</sup> T cell co-culture system to functionally test for immunosuppressive properties of sepsis neutrophils (Fig. 5.11B). I found that when culturing neutrophils with CD4<sup>+</sup> T cells stimulated with anti CD3/28 beads, the cultures with sepsis neutrophils significantly inhibited CD4<sup>+</sup> T cell proliferation (measured by proliferative fraction) and activation (measured by PD-1 and CD69 expression) (Fig. 5.11C), confirming sepsis neutrophils as bona fide MDSCs.

Increased arginase-1 activity and PD-L1/2 inhibition of T cells have been proposed as mechanisms for neutrophil immunosuppression (Veglia et al. 2018), but I was unable to reverse the suppressive effects with either anti-PDL1/2 antibodies or arginase-1 inhibitors (Fig. 5.12). Instead, I was able to partially reverse suppression of T cell activation by inhibiting the prostaglandin EP2 receptor, as reflected by partial restoration of CD4<sup>+</sup> T cell CD69 expression in sepsis neutrophil co-cultures when adding the EP2 receptor inhibitor TG6-10-1 (Fig. 5.12).

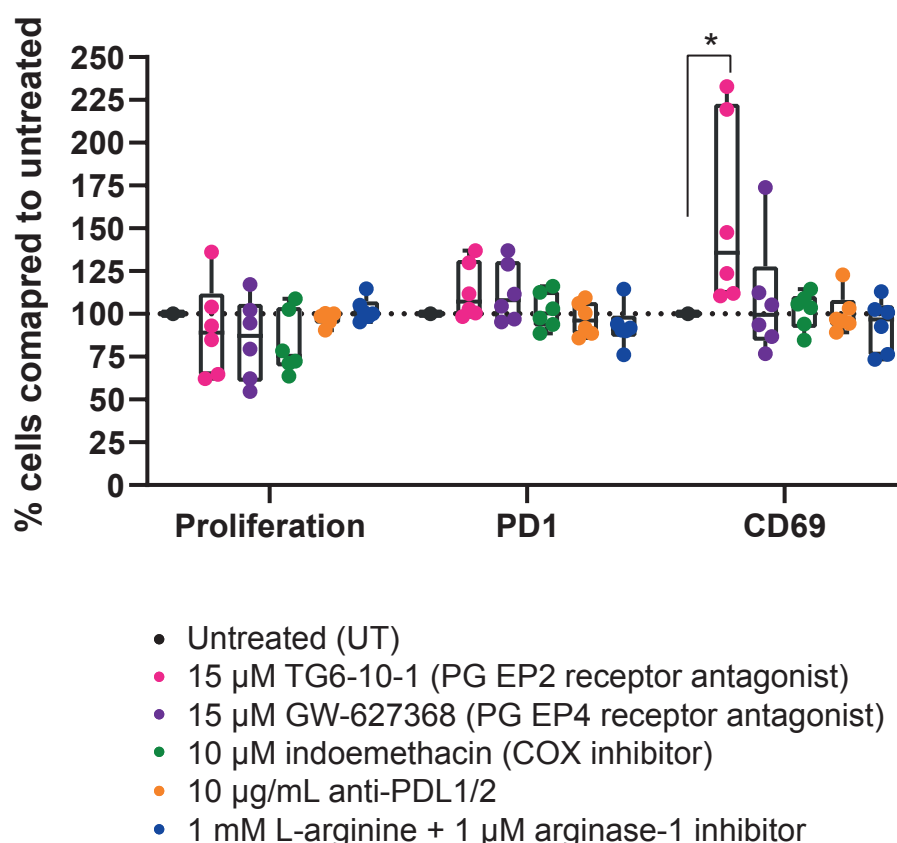
Neither addition of indomethacin (cyclo-oxygenase inhibitor) nor inhibition of the prostaglandin EP4 receptor showed a similar effect, suggesting a mechanism





**Figure 5.11: Sepsis neutrophil immunosuppressive properties.** (A) Gene set scoring of cohort 1 single-cell RNA/protein profiling mature and immature neutrophils in sepsis and healthy volunteer (HV) samples by gene signatures of granulocytic myeloid derived suppressor cells. (B) Schematic of neutrophil-CD4<sup>+</sup> T cell co-culture. (C) Proliferation (top), and activation as measured by surface PD-1 (middle) and CD69 (bottom) positive anti-CD3/28 bead stimulated CD4<sup>+</sup> T cells after co-culture with neutrophils at a 4 neutrophil:1 T cell ratio, compared to positive controls of CD4<sup>+</sup> T cells cultured with anti-CD3/28 beads alone. HV n=10, sepsis n=17 for proliferation, n=16 for PD-1 and n=14 for CD69. All P-values were calculated with two-sided Wilcoxon rank-sum tests. \*P < 0.05, \*\*P < 0.01, \*\*\*P < 0.001, \*\*\*\*P < 0.0001.

of action at the level of pre-formed prostaglandin E2 acting at the EP2 receptor. Moreover, I was only able to partially restore CD69 expression and could not modulate PD-1 expression or proliferation suppression, suggesting that other mechanisms explain the neutrophil mediated CD4<sup>+</sup> T cell suppression.



**Figure 5.12: Reversal of neutrophil mediated CD4<sup>+</sup> T cell suppression.** Percentage of CD4<sup>+</sup> T cells proliferating, expressing PD-1 or expressing CD69 in sepsis neutrophil-allogeneic T cell co-cultures with various treatments compared to UT control (n=6). All comparisons not significant except for CD69 for TG6-10-1 vs. UT. *P*-values were calculated with two-sided Wilcoxon rank-sum tests. \**P* < 0.05.

### 5.2.6 Understanding cellular compositional drivers of SRS

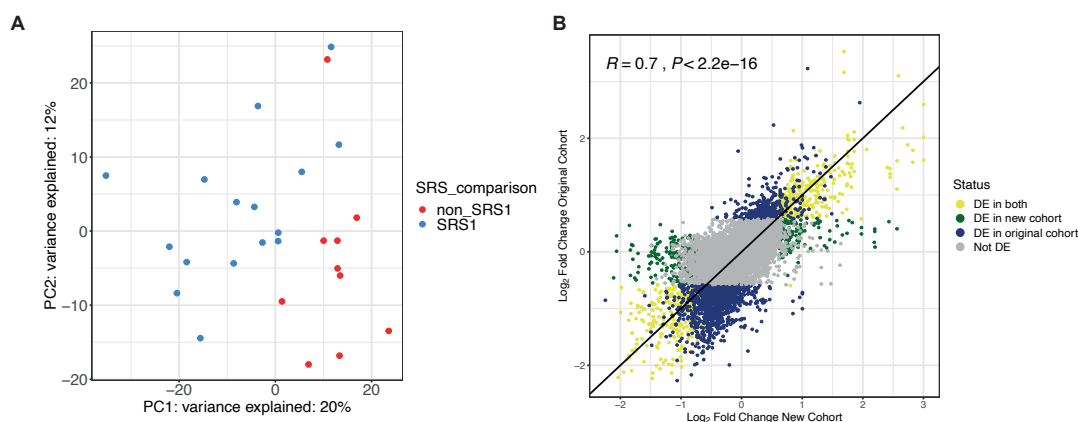
Beyond defining the septic state from sterile inflammatory or healthy conditions, one major clinical issue surrounding sepsis involves the difficulty in resolving the heterogeneity of the condition to stratify patients for meaningful prognostication

and rational approaches to therapeutic trials. To this end, I assigned to our cohort the status of Sepsis Response Signatures (SRS), defined previously via peripheral blood bulk transcriptomics and robustly replicated across cohorts and causes of sepsis to predict early mortality (Davenport et al. 2016; Burnham et al. 2017; Cano-Gamez et al. 2022). SRS1 patients or patients with high SRSq score have a higher mortality and display features of immunosuppression.

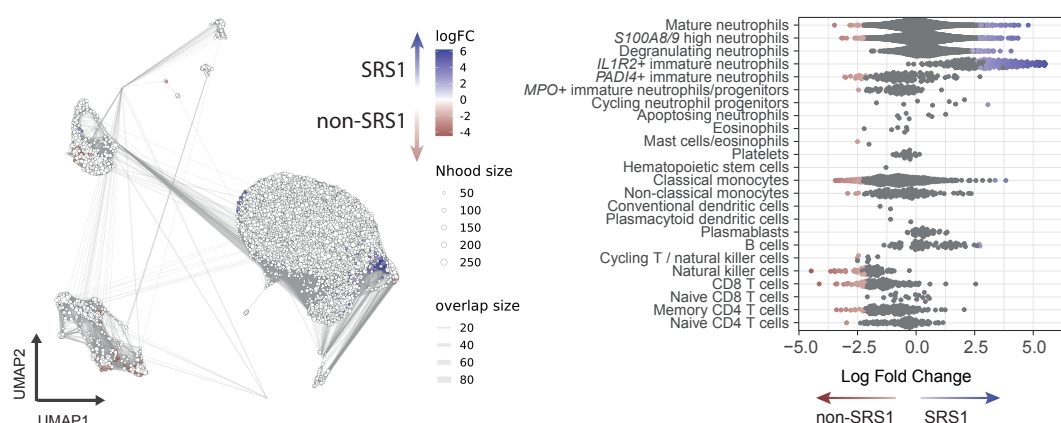
I assigned SRS status and SRSq scores for the 26 acute sepsis samples using the previously defined seven gene set (*DYRK2*, *CCNB1IP1*, *TDRD9*, *ZAP70*, *ARL14EP*, *MDC1*, and *ADGRE3*) and the algorithm implemented in the SepstratifieR package (Cano-Gamez et al. 2022). Based on pseudobulk profiles of the whole blood single-cell dataset, I identified 16 SRS1 samples and 10 non-SRS1 samples. Principal components analysis (PCA) of the top 15% most variable genes of our pseudobulk profiles showed separation of the SRS groups along the first two principal components (PCs) (Fig. 5.13A). The robustness of the SRS assignment was assessed by correlating the fold changes in gene expression between SRS1 vs. non-SRS1 samples with the fold changes in the cohort from which SRS was originally derived (Davenport et al. 2016). I found a strong and highly significant correlation, supporting the validity of SRS signals in this cohort (Fig. 5.13B).

Testing for differential abundance across SRS groups revealed an expansion of IL1R2+ immature neutrophils in SRS1 compared to non-SRS1 patients, as well as minor enrichments in *S100A8/9* high neutrophils, degranulating neutrophils and cycling neutrophil progenitors, with NK cell, memory CD4+ and CD8+ T cell and cMono depleted in SRS1 (Fig. 5.14).

To understand the cellular gene expression basis of SRS, I performed pseudobulk differential gene expression (DGE) analysis testing against SRS as the covariate of interest, accounting for age, sex and sequencing batch. A large number of genes were differentially expressed at a log2FC of 1.2 and FDR <0.1 for mature



**Figure 5.13: SRS signatures in cohort 1 single cell dataset.** (A) First two PCs from PCA of pseudobulked single cell expression profiles of sepsis samples (all cells per sample) ( $n=16$  SRS1,  $n=10$  non-SRS1). (B) Correlation of fold change in gene expression between SRS1 vs. non-SRS1 in cohort 1 and cohort in which SRS was originally derived (Davenport et al. 2016).



**Figure 5.14: Cohort 1 cellular composition analysis across SRS.** Differential abundance (DA) analysis using graph neighborhood (Nhood) based method to detect enrichment of neighborhoods across SRS state within sepsis samples ( $n=16$  SRS1,  $n=10$  non-SRS1), with UMAP of sampled neighborhoods and statistically significant enrichment colored (spatial FDR  $< 0.05$  with generalized linear modeling). (B) Beeswarm plot of SRS1 vs. non-SRS1 DA analysis with cluster labels of neighborhoods depicted.

neutrophils, *S100A8/9* high neutrophils, degranulating neutrophils and *IL1R2*+ immature neutrophils which translated into separation of samples by SRS status on PCA (Fig. 5.15A-B), consistent with the importance of neutrophils in the SRS signal. Strikingly, I observed minimal differential gene expression (DGE) in any of the MNC subsets, suggesting a minor and passive contribution of MNCs to SRS definitions (Fig. 5.16).

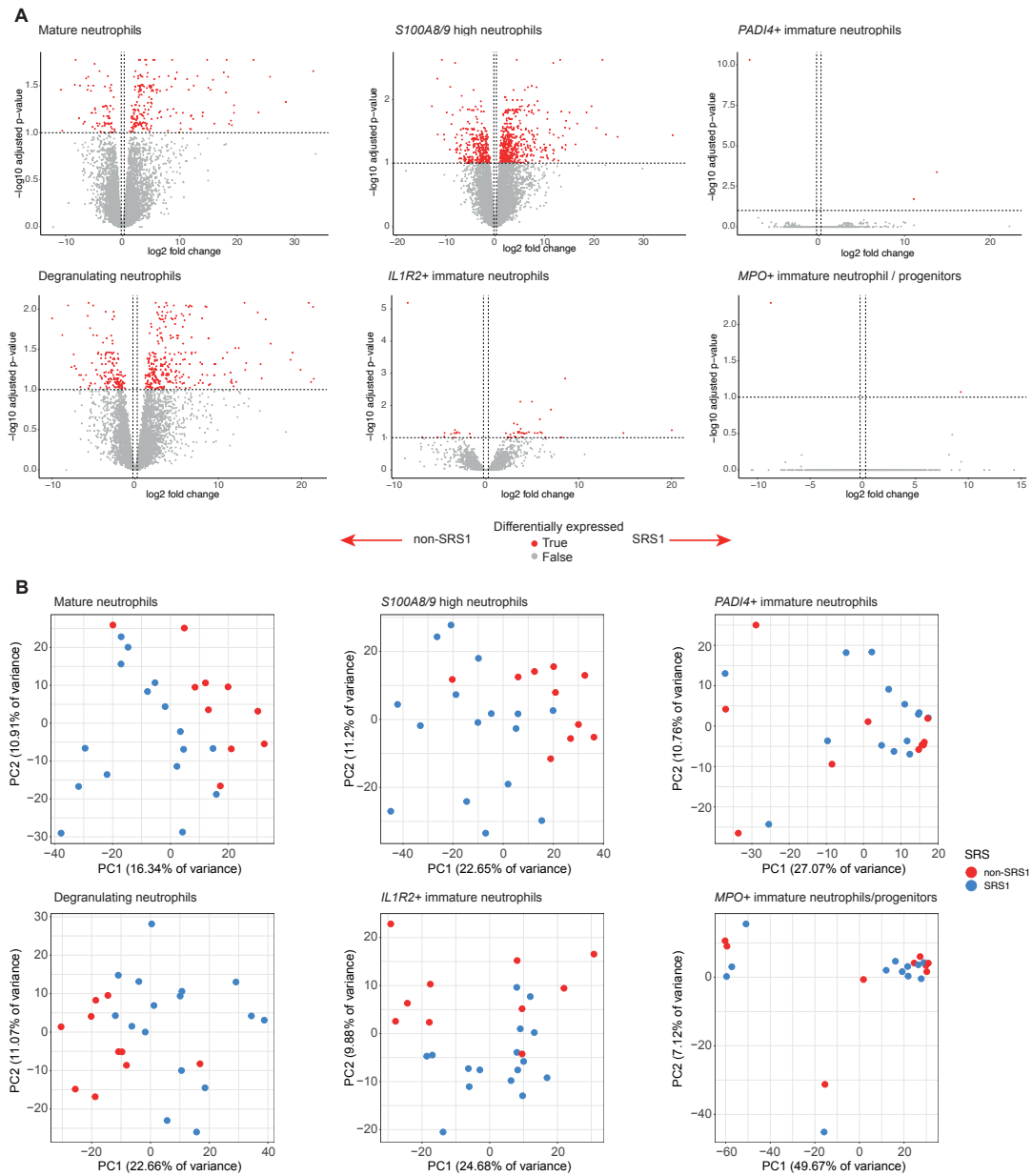
### 5.2.7 Neutrophil function in SRS

To understand whether these neutrophil changes may be related to sepsis pathogenesis, I analyzed the neutrophil T cell co-culture data in the setting of SRS. Functionally, there was no difference across SRS in neutrophil suppression of CD4+ T cell proliferation (Fig. 5.17A), but I found that SRS1 neutrophils suppressed CD4+ T cell activation more than non-SRS1 neutrophils as measured by T cell CD69 and PD1 expression (Fig. 5.17B). Moreover, SRS1 neutrophils displayed reduced phagocytosis, which I was unable to restore with GM-CSF (Fig. 5.17C). Taken together, the data were consistent with the hypothesis that SRS1 represents an immunosuppressed state.

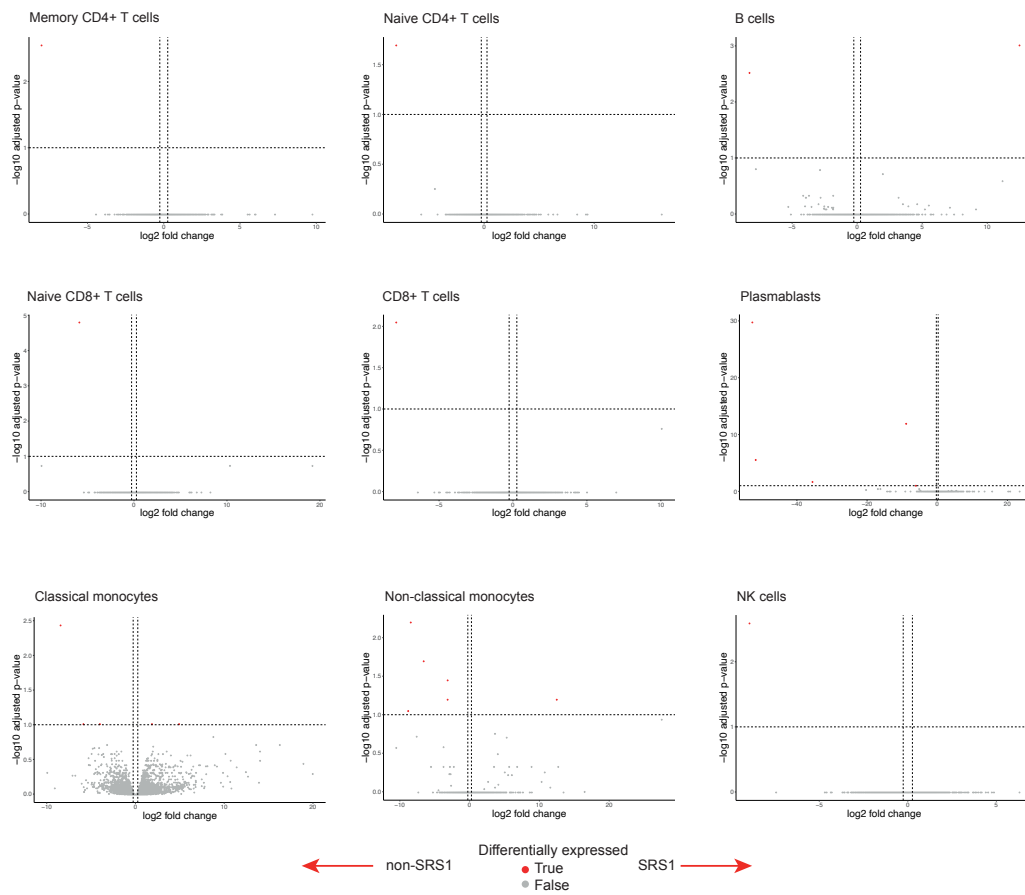
### 5.2.8 Validating neutrophil contribution to SRS with orthogonal datasets

To test the external validity of the results, I first sought evidence of increased immature neutrophils in SRS1. I re-analyzed whole blood bulk RNA-seq and mass cytometric (CyTOF) immunophenotyping data from the 36 all-cause sepsis patients (42 samples) from COMBAT (cohort 2) by SRS (Fig. 5.1).

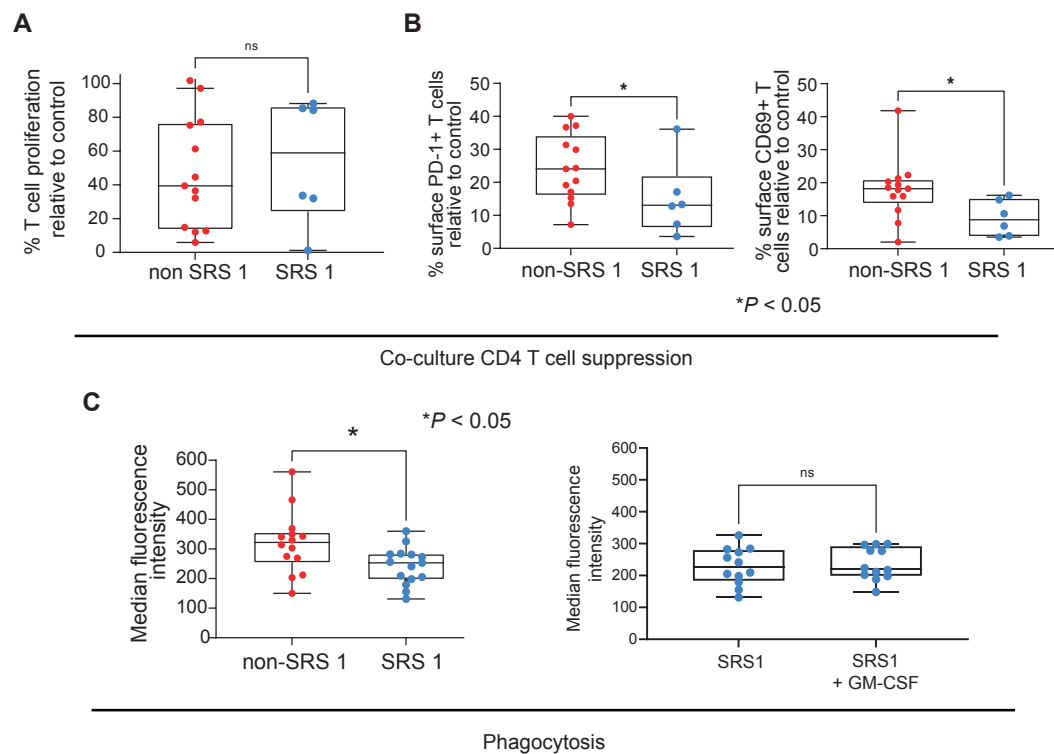
I reduced dimensions of the RNA-seq dataset with weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath 2008) (chapter 2). Of the 33 distinct gene modules extracted, differential eigengene expression testing yielded 8 upregulated modules in SRS1 and 5 downregulated (Wilcoxon rank



**Figure 5.15: Cohort 1 SRS neutrophil pseudobulk gene expression analysis.** (A) Volcano plots of DGE analysis between SRS groups (n=16 SRS1, n=10 non-SRS1) for pseudobulked neutrophil states (red denoting genes with fold change  $\geq 1.2$  and FDR  $< 0.1$ ) (positive fold change denoting upregulation in SRS1). (B) First two PCs from PCAs of pseudobulked neutrophil states colored by SRS groups.



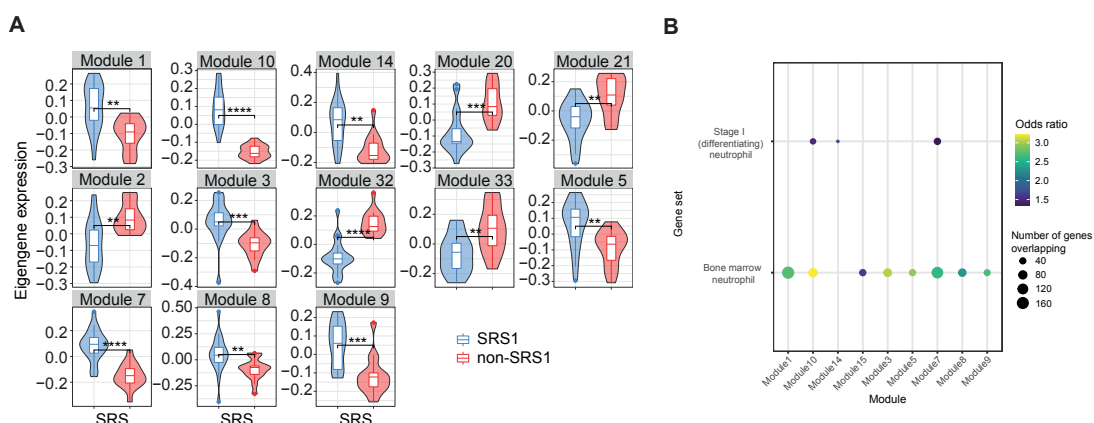
**Figure 5.16: Cohort 1 SRS mononuclear cell pseudobulk gene expression analysis.** Volcano plots of DGE analysis between SRS groups (n=16 SRS1, n=10 non-SRS1) for pseudobulked mononuclear cells (red denoting genes with fold change  $\geq 1.2$  and FDR  $< 0.1$ ) (positive fold change denoting upregulation in SRS1).



**Figure 5.17: Functional interrogation of neutrophils across SRS.** Co-culture done with sepsis neutrophils at a 4 neutrophil:1 T cell ratio. (A) Proliferation (left), and activation as measured by surface PD-1 (middle) and CD69 (right) positive anti-CD3/28 bead stimulated CD4+ T cells after co-culture, compared to positive controls of CD4+ T cells cultured with anti-CD3/28 beads alone. SRS1 n=6, non-SRS1 n=13 (B) Neutrophil phagocytosis across SRS. SRS1 n=15, non-SRS1 n=14 (C) Neutrophil phagocytosis with and without GM-CSF stimulation (n=12). *P*-values were calculated with two-sided Wilcoxon rank-sum tests. \**P* < 0.05, ns = not significant.



sum test FDR <0.01) (Fig. 5.18A). Overlapping the module genes with signatures of differentiating and differentiated neutrophils from mouse single cell transcriptomics examining hematopoietic trajectories (Giladi et al. 2018), I found that modules 10, 14 and 7 were significantly enriched for differentiating neutrophils (Fig. 5.18B).

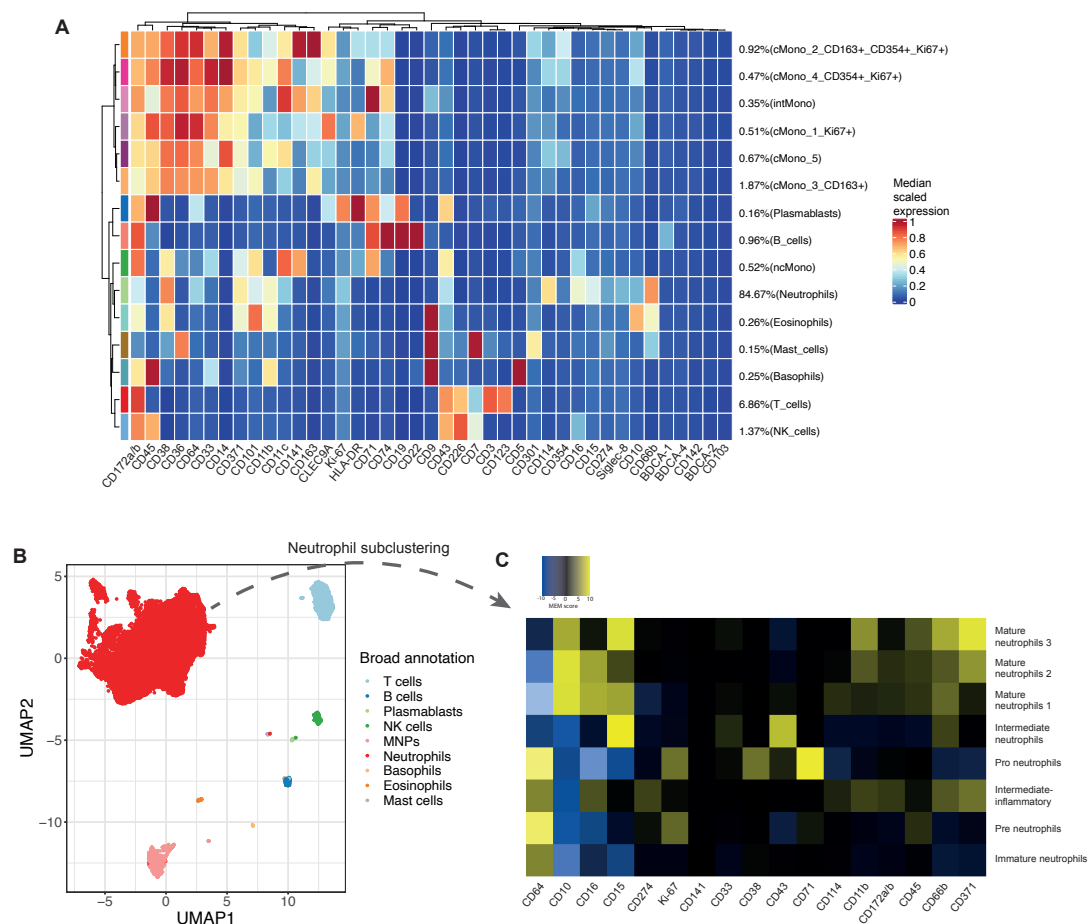


**Figure 5.18: Cohort 2 bulk RNA-seq weighted gene network correlation analysis (WGCNA).** (A) WGCNA gene modules with differential eigengene expression across SRS in whole blood bulk RNA-seq (FDR <0.01, two-sided Wilcoxon rank-sum test). (B) Enrichment of neutrophil related gene sets (bottom, human cell atlas bone marrow cell type gene sets (Hay et al. 2018); top, gene sets of differentiating neutrophils from mouse single-cell transcriptomic data (Giladi et al. 2018)) of WGCNA modules upregulated in SRS1.

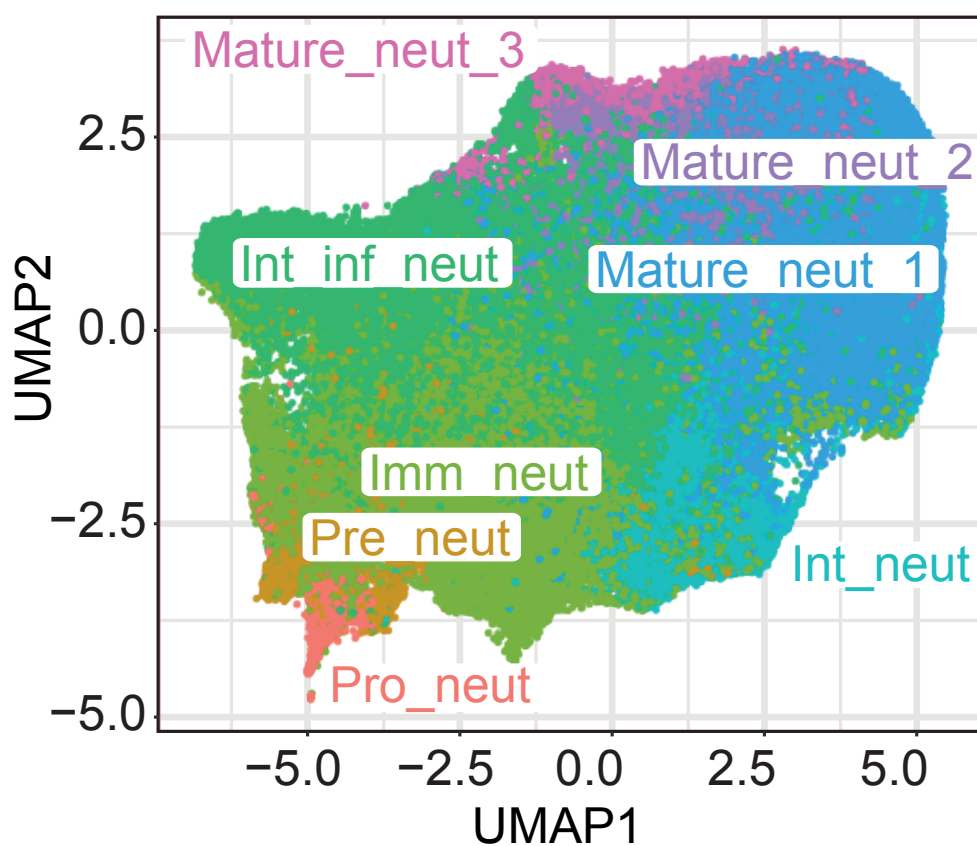
To elucidate the cellular basis of these signals in SRS1, I re-analyzed CyTOF data available for 41 of the 42 sepsis samples in COMBAT and 11 age and sex matched healthy controls by iterative clustering, merging and annotating cell types and states (chapter 2, Fig. 5.19A).

I subclustered the neutrophils with 17 selected markers, assigning them to eight distinctively different states according to Marker Enrichment Modeling (Diggins et al. 2017) which spanned a range of maturity stages (chapter 2; Fig. 5.19C; Fig. 5.20). On mixed model differential abundance testing, I found more immature and pro/pre-neutrophils in SRS1 (Fig. 5.21A-B).

To analytically confirm the enrichment of immature neutrophils in SRS1, I performed pseudotime trajectory analysis with diffusion map dimensionality

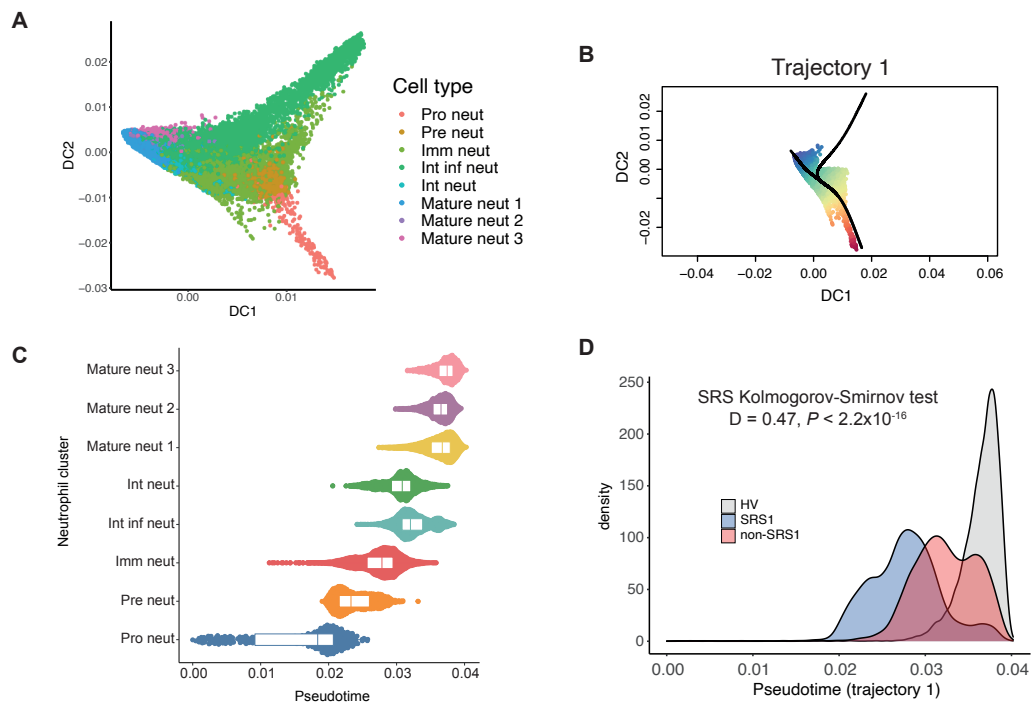


**Figure 5.19: Cohort 2 CyTOF neutrophil cluster identification.** (A) Heatmap of 15 immune cell clusters (with neutrophil clusters merged; Methods) from clustering on 41 protein markers for  $n = 41$  sepsis (36 individuals) and 11 healthy control samples. (B) UMAP of nine broad immune cell types (2,287,410 cells) from CyTOF immunophenotyping. (C) Marker enrichment modeling (MEM) score heatmap of eight neutrophil clusters (1,921,471 cells), identified from neutrophil subclustering on 17 selected protein markers. cMono = classical monocyte, ncMono = non-classical monocyte, NK cell = natural killer cell.



**Figure 5.20: Cohort 2 CyTOF neutrophil cluster UMAP visualization.** CyTOF neutrophil subsets (1,921,471 cells) from healthy volunteers (n=11) and sepsis samples (n=41).





**Figure 5.22: Cohort 2 CyTOF neutrophil trajectory analysis.** (A) Diffusion map dimensionality reduction of neutrophils (45,000 cells, 15,000 from each comparator group i.e. healthy volunteers (HV), SRS1 and non-SRS1). (B) Diffusion components 1 and 2 with black line depicting trajectory identified by principal curve fitting. Color reflects pseudotime value of cell with red indicating early pseudotime and blue late pseudotime. (C) Density distribution of neutrophils over pseudotime with Kolmogorov-Smirnov test for SRS1 vs. non-SRS1 distributions.

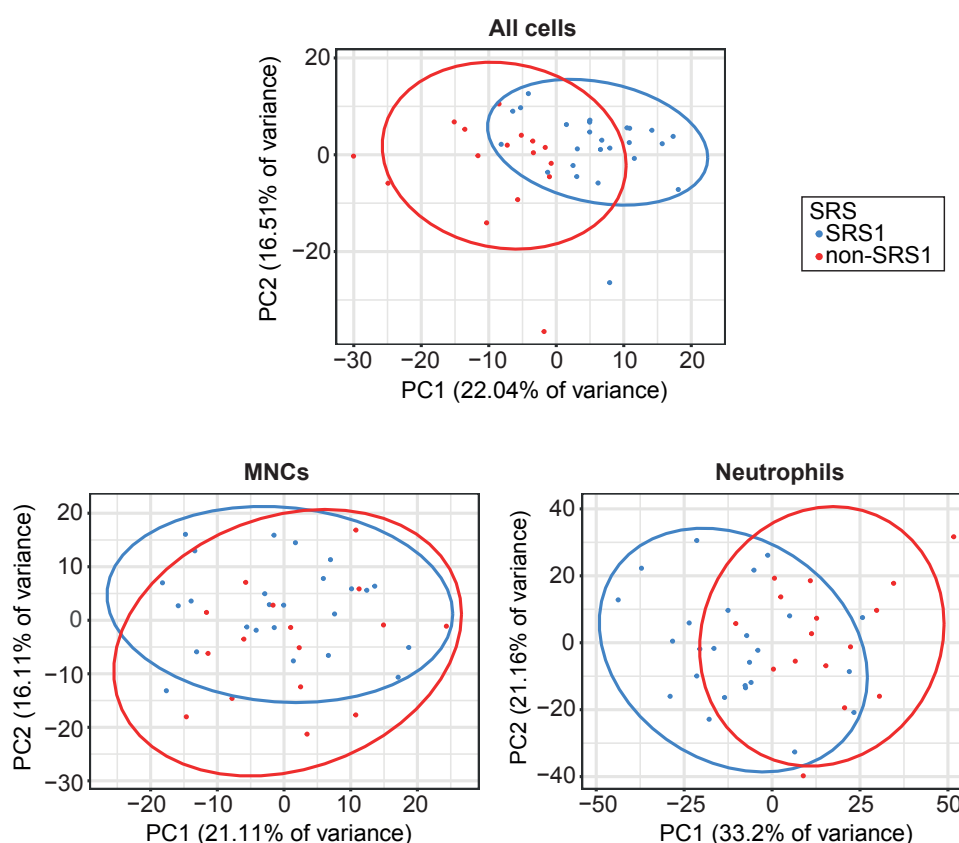
reduction and principal curve fit (Street et al. 2018). Diffusion maps were invented to reduce dimensionality of datasets while retaining the original connectivity between data points (Coifman and Lafon 2006). This analysis revealed a trajectory which ordered cells along pseudotime (Fig. 5.22A-C). I found that the density of neutrophils in the earlier and more immature stages of the trajectory was significantly greater in SRS1 than non-SRS1 (Fig. 5.22D, Kolmogorov-Smirnov test  $D=0.47$ ,  $P < 2.2 \times 10^{-16}$ ).

To understand the differential contributions of neutrophils and MNCs to SRS, I performed PCA on cell proportions within the neutrophil and MNC compartments. This showed that, in contrast to neutrophils, proportional differences in MNCs in isolation could not separate the SRS groups, confirming the centrality of neutrophils to SRS definitions (Fig. 5.23).

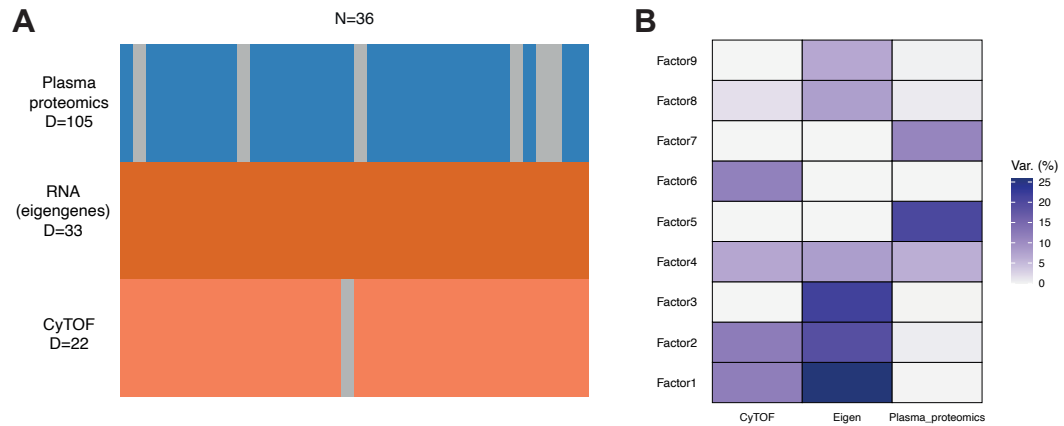
### **5.2.9 CyTOF and bulk RNA-seq integration by factor analysis**

As SRS definitions are derived from gene expression data, I next sought to integrate cohort 2 CyTOF cell composition data with bulk RNA-seq data with factor analysis (Argelaguet et al. 2020). By identifying the latent spaces containing shared variation across the data modalities, I could connect the gene expression signatures underlying SRS groups with cell clusters, thus directly implicating specific cell clusters in driving SRS membership. I included 105 plasma protein measurements from the same patients as a computational negative control, reasoning that the latent spaces linking cell clusters to gene expression profiles should not show an overly strong contribution from the plasma proteomics as the plasma proteome reflects contributions from various tissues (Anderson and Anderson 2002) while the RNA was derived directly from leukocytes (Fig. 5.24A-B).

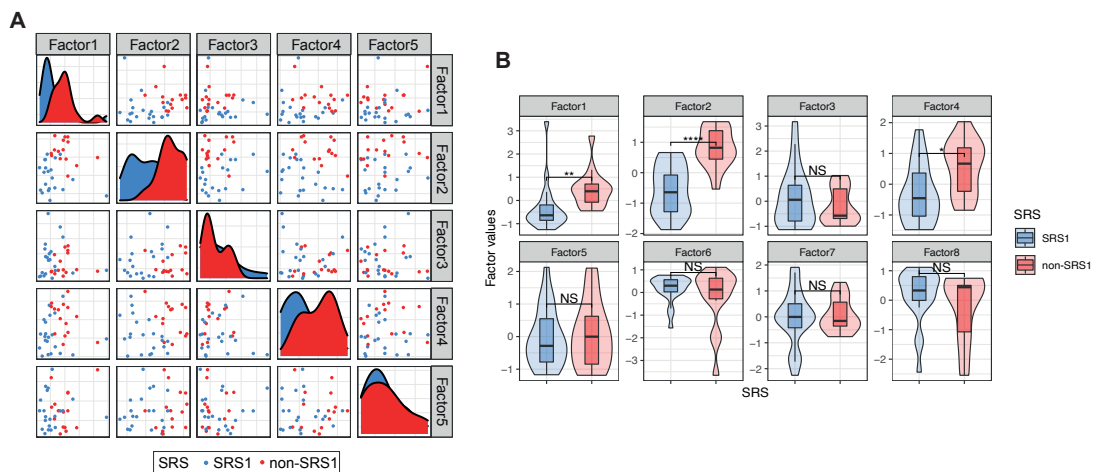
Factors one and two were responsible for most of the observed SRS separation (Fig. 5.25A) and were significantly different between SRS groups (Fig. 5.25B).



**Figure 5.23: Cohort 2 CyTOF principal components analysis (PCA) on cell proportions.** First two principal components of PCAs on all cell proportions (top), neutrophils only (bottom left) and mononuclear cells only (bottom right), with 95% data ellipses (assuming a multivariate t-distribution). SRS1 n=25, non-SRS1 n=16



**Figure 5.24: Cohort 2 multi-omic factor analysis plus (MOFA+) overview.** (A) Input data into MOFA+ model consisting of three modalities (105 plasma proteins, 33 module eigengenes from WGCNA and 22 cell clusters from CyTOF) for n=36 sepsis samples with one and six missing samples for CyTOF and plasma proteomics respectively. D = dimensionality of data modality (B) Variance decomposition showing percentage variance explained for individual modalities by each latent factor.



**Figure 5.25: Cohort 2 multi-omic factor analysis plus (MOFA+) by SRS.** (A) Scatter plots of latent factor values for sepsis patients (n=36). (B) Factor values by SRS groups (two-sided Wilcoxon rank sum-test). \*FDR <0.05, \*\*FDR <0.01, \*\*\* FDR <0.001, \*\*\*\* FDR <0.0001.



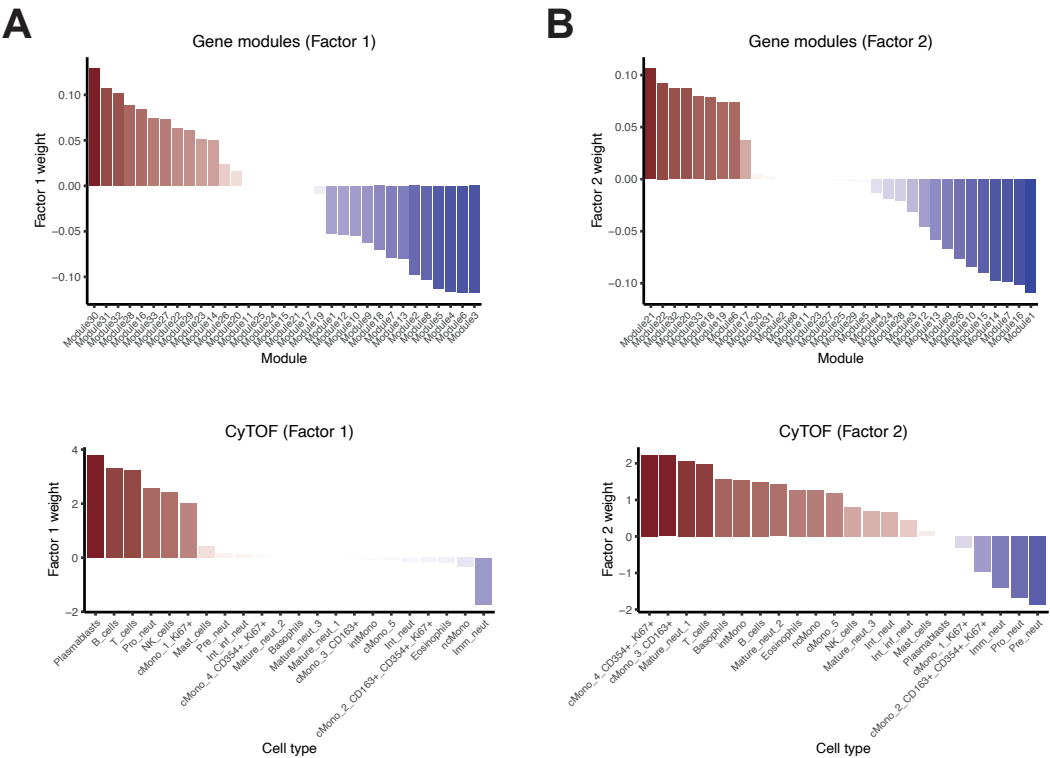
These two factors explained minimal variance within the plasma proteomic dataset (Fig. 5.24B). By contrast, factor five, which accounted for most of the variance in the plasma proteomic data (20.3%), was not different across SRS and accounted for <0.01% of variance in the gene expression and CyTOF data (Fig. 5.24B; Fig. 5.24B). This suggested that differences in plasma protein levels did not contribute substantially to SRS, while the genes and cell types separating SRS1 and 2 along the first two factors were genuine drivers of SRS.

Of the 13 WGCNA gene modules driving a low factor one value (in the direction of SRS1), six (modules 1, 3, 8, 7, 9, and 10) were significantly upregulated in SRS1 in the differential co-expression analysis (Fig. 5.18A). Conversely, two of seven modules (modules 32, 33) contributing to a high factor one value (in the direction of non-SRS1) were significantly upregulated in non-SRS1 (Fig. 5.26A; Fig. 5.18A). The key cell type contributing to the SRS1 direction was immature neutrophils, while the various MNCs (T, B and NK cells and plasmablasts) contributed in the opposite direction (Fig. 5.26A).

Similarly, five of the top eight modules contributing to a negative factor two value in the direction of SRS1 (modules 1, 7, 14, 10 and 9) were upregulated in SRS1, while four of the five top modules leading to a positive factor two value had been significantly upregulated in non-SRS1 (Fig. 5.26B; Fig. 5.18A). Correspondingly, I found strong negative weighting in the CyTOF modality for factor two to include pre-neutrophils, pro-neutrophils and immature neutrophils (Fig. 5.26B), highlighting the relevance of these immature neutrophil populations to the SRS1 state.

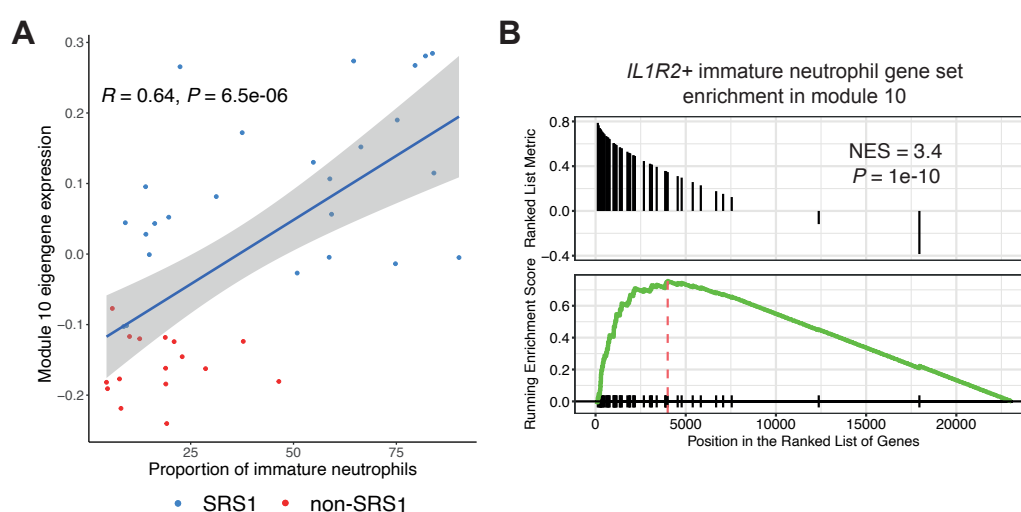
### **5.2.10 Validating expansion of *IL1R2*<sup>+</sup> immature neutrophils**

While the definition of neutrophil states was based on different features in cohort 1 and cohort 2, I nevertheless aimed to test for enrichment of the specific *IL1R2*<sup>+</sup> immature neutrophil population that I had specifically isolated



**Figure 5.26: Cohort 2 multi-omic factor analysis plus (MOFA+) factor loadings.** (A-B) Weights per cell cluster and gene module for (A) latent factor one and (B) latent factor two.

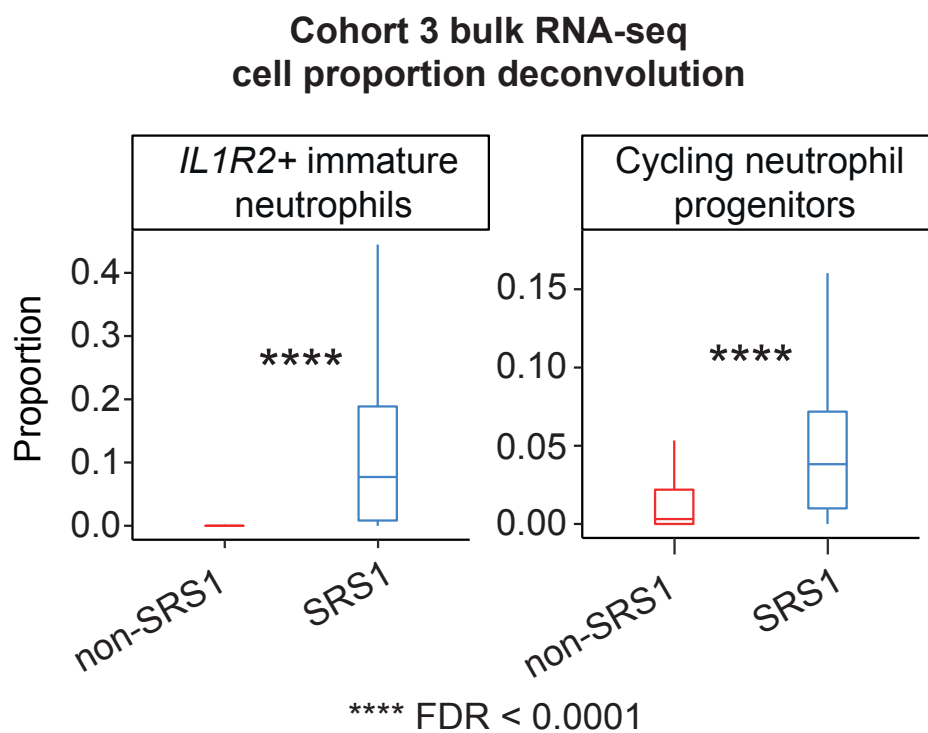
in cohort 1 in the validation dataset. I first asked which gene module was most correlated with the immature neutrophils as defined by CyTOF, finding this to be module 10 (Pearson's  $r=0.64$ ,  $P=6.5 \times 10^{-6}$ ) (Fig. 5.27A). Leading edge analysis of module 10 genes ranked by correlation with eigengene expression indeed showed enrichment of a signature defining IL1R2+ immature neutrophils (Fig. 5.27B), providing direct evidence of the concordance of the analysis from the two datasets.



**Figure 5.27: Cohort 2 *IL1R2*+ immature neutrophil validation.** (A) Correlation of module 10 eigengene expression against proportion of immature neutrophils from mass cytometry (CyTOF) per sample of cohort 2 ( $n=41$ , 25 SRS1 and 16 non-SRS1). (B) Leading edge analysis for enrichment of cohort 1 derived *IL1R2*+ immature neutrophil defining signature in WGCNA module 10.

To expand beyond the samples analyzed which contained both immune cell and gene expression data, I performed cell type/state proportion deconvolution with CIBERSORTx (Newman et al. 2019) on a further 864 SRS assigned sepsis samples (EGAD00001008730) for which only bulk whole blood gene expression data was available (cohort 3) by leveraging the single cell multi-omic dataset to build a reference.

This cohort consisted of adult sepsis patients recruited from the intensive care unit (ICU) with samples collected within 5 days of ICU admission. Differential abundances of cell populations predicted from deconvolution for the 667 first



**Figure 5.28: Cohort 3 *IL1R2+* immature neutrophil validation.** Cell type/state proportions in cohort 3 sepsis patients (n=667) following whole blood bulk RNA-seq deconvolution (two-sided Wilcoxon rank-sum test). \*\*\*\*FDR <0.0001.

available samples per patient showed significant *IL1R2*<sup>+</sup> immature neutrophil and cycling neutrophil progenitor expansion in SRS1 using univariate tests (Fig. 5.28), recapitulating the observations from the multi-modal analyses.

### 5.2.11 Understanding the mediators and pathways that drive SRS

As genes function in concert in biological pathways, I further interrogated gene expression in various neutrophil states from cohort 1 samples with consensus non-negative matrix factorisation (Methods) (Kotliar et al. 2019) to characterize differential activity across SRS at the level of gene programs rather than of individual genes (Fig. 5.29). The motivation for the use of this technique was similar to that of using WGCNA in the previous chapter.

At FDR <0.05, program activity was correlated with SRSq scores in three programs in mature neutrophils, *S100A8/9* high neutrophils and *IL1R2*<sup>+</sup> immature neutrophils, and in two programs in *PADI4*<sup>+</sup> immature neutrophils (Fig. 5.30A). Pathway enrichment testing of programs positively correlated with SRSq against Hallmark gene sets from MsigDB revealed enrichment in IL-6 JAK-STAT3 signaling, TNF, MTORC1, hypoxia and complement (Fig. 5.30B).

GO over-representation enrichment testing of the top 50 genes of *S100A8/9* high neutrophil program 2 (negative correlation with SRSq) showed response to glucocorticoids (Fig. 5.30C), notable given the interaction between steroid administration, SRS and mortality in sepsis (Antcliffe et al. 2019). For program 8 of mature neutrophils, GO pathway analysis highlighted response to IL-1 for mature neutrophils (Fig. 5.30D), relevant given the expansion of *IL1R2*<sup>+</sup> neutrophils, as *IL1R2* is a decoy receptor for IL-1 $\beta$ . The glycolytic pathway was also highlighted in *IL1R2*<sup>+</sup> immature neutrophils (Fig. 5.30E), revealing an additional cell state which displays altered glycolytic metabolism in sepsis in addition to monocytes (Cheng et al. 2016).

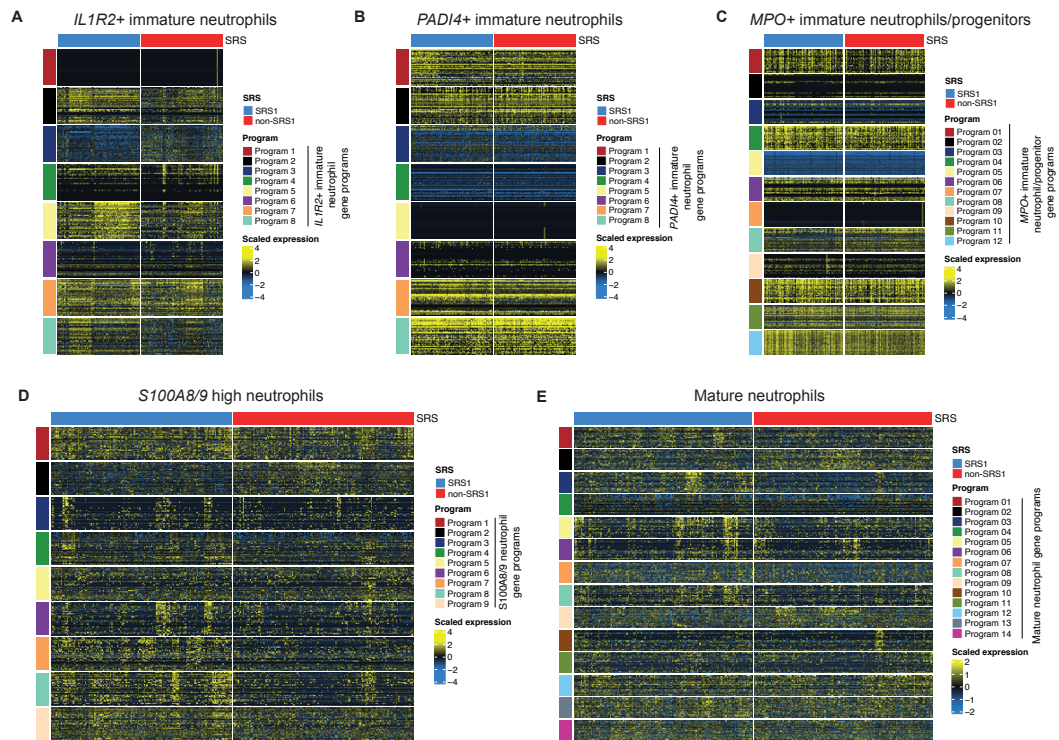
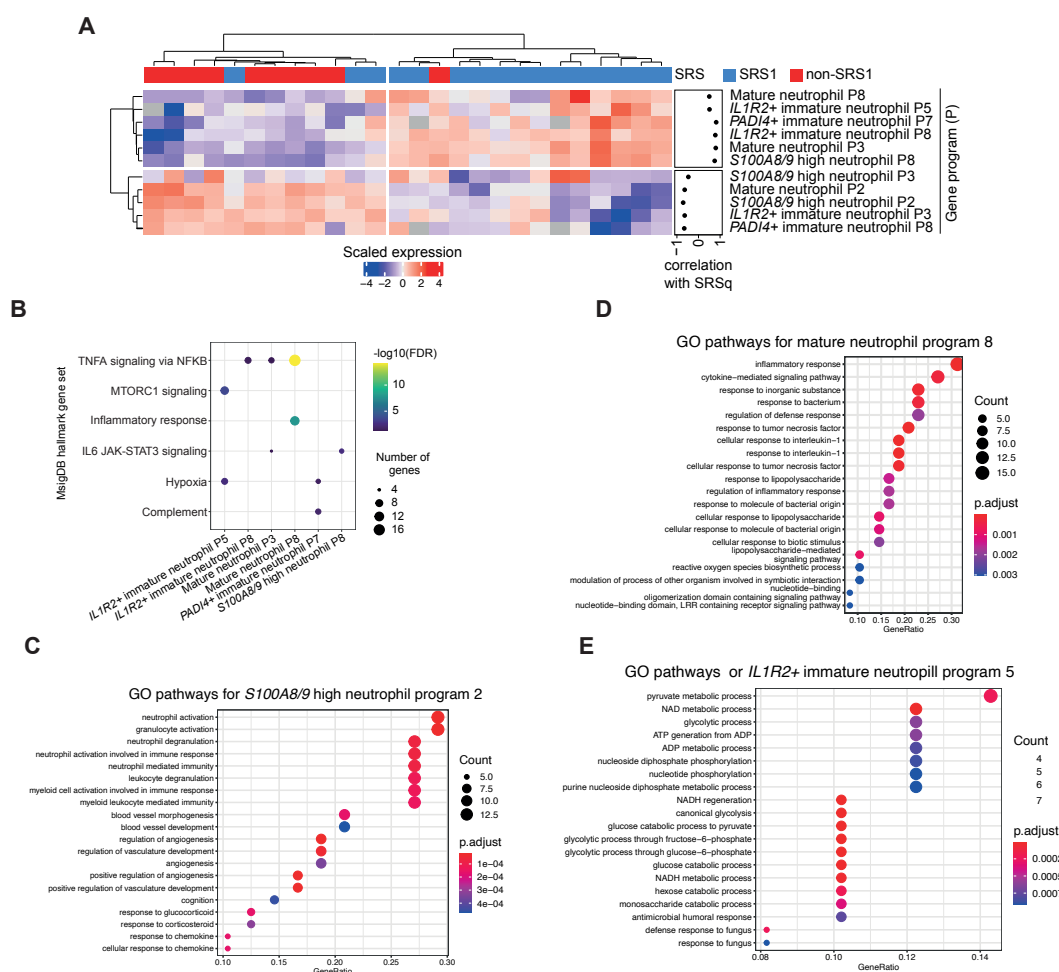


Figure S12

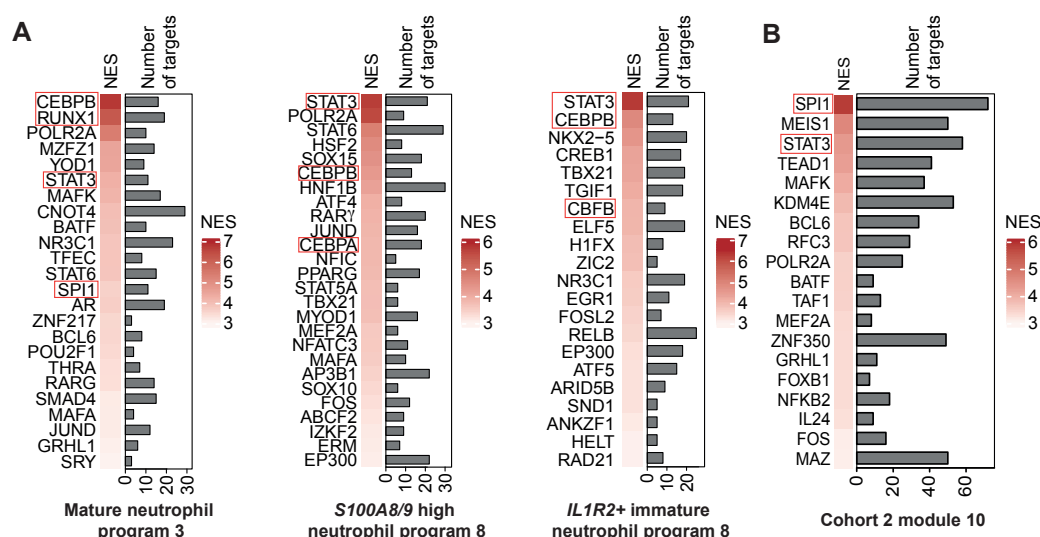
**Figure 5.29: Cohort 1 SRS neutrophil consensus non-negative matrix factorization (cNMF) analysis.** (A-E) Heatmaps of gene programs from cNMF of cohort 1 (A) *IL1R2*+ immature neutrophils (B) *PADI4*+ immature neutrophils (C) *MPO*+ immature neutrophils/progenitors (D) *S100A8/9* high neutrophils and (E) mature neutrophils. Each heatmap shows 150 downsampled SRS1 and non-SRS1 cells. Each neutrophil state was analyzed independently i.e. programs are not shared between each neutrophil state.



**Figure 5.30: Gene programs in cohort 1 neutrophils** (A) Gene expression heatmap for cohort 1 neutrophil programs significantly correlated with SRSq (Spearman  $Rho$ , FDR  $< 0.05$ ). (B) MsigDB Hallmark gene set over-representation tests pathway enrichment for programs with significant positive correlations with SRSq. (C-E) Over-representation test pathway enrichment dotplots of top Gene Ontology (GO) pathways for gene programs of (C) *S100A8/9* high neutrophils (program 2) (D) mature neutrophils (program 8) and (E) *IL1R2+* immature neutrophils (program 5).

# Transcription factors driving SRS neutrophil programs

In view of the enrichment of JAK-STAT3 signaling (Fig. 5.30B) in two neutrophil gene programs correlated to high SRSq scores, I proceeded to perform transcription factor (TF) prediction analysis for the gene programs with higher expression in SRS1. Using the iRegulon method which detects enriched TF motifs in gene sets based on an extensive collection of TF position weight matrices from multiple species (Janky et al. 2014), I identified a number of TFs relevant to myelopoiesis in programs 3, 8 and 8 in mature, *S100A8/9* high and *IL1R2*+ immature neutrophils respectively. This included RUNX1, SPI1 and CEBFB, CEBPA, notable for its role in steady state granulopoiesis (SSG) and CEBPB, a key TF for emergency granulopoiesis (EG) (Manz and Boettcher 2014). There was also enrichment for STAT3, another TF central to EG (Zhang et al. 2010), in all three programs (Fig. 5.31A).



**Figure 5.31: Cohort 1 transcription factor (TF) predictions.** (A) TF prediction analysis (top 50 genes per program) for three cohort 1 neutrophil programs positively correlated with SRSq. (B) TF prediction analysis for top 1% genes showing highest correlation with whole blood bulk RNA-seq module 10 eigengene.

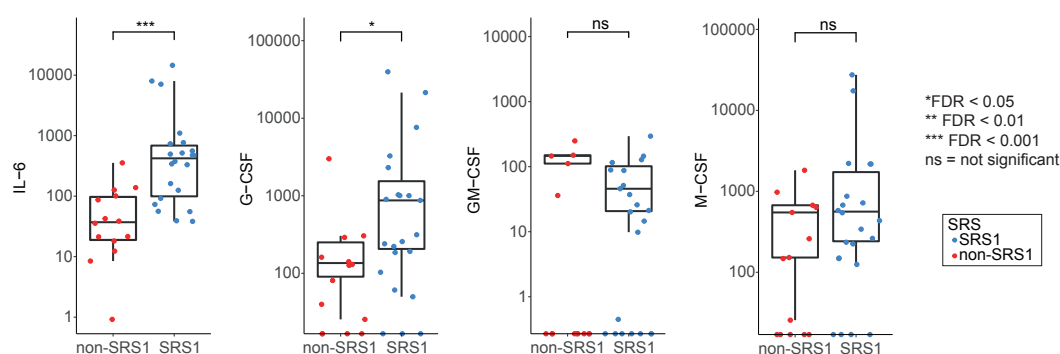
I found further evidence for myelopoiesis and enhanced STAT-3 signaling in cohort 2 via TF prediction analysis for gene expression module 10 (module most



strongly correlated with immature neutrophil proportion) which revealed SPI1 and STAT-3 as predicted regulators (Fig. 5.31B).

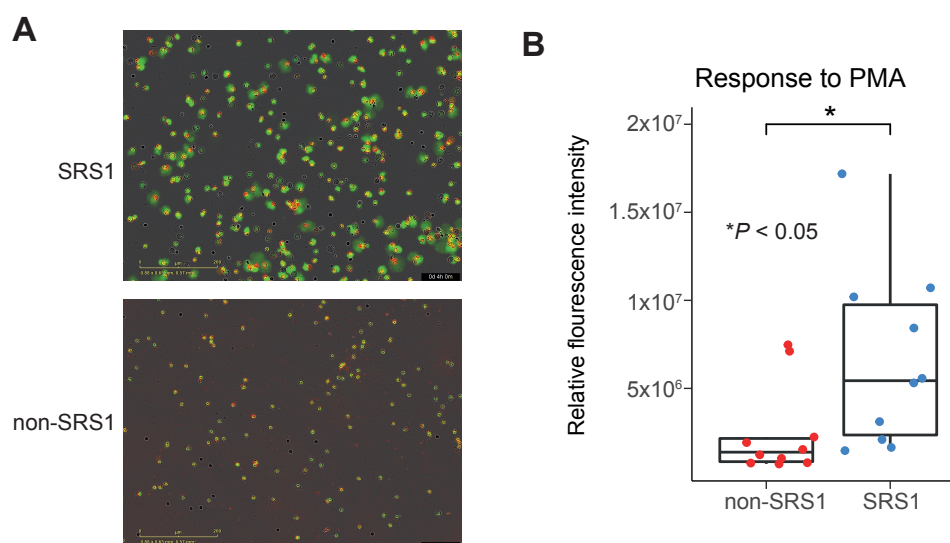
### Unraveling the STAT3 signals

Given the myelopoietic and STAT3 TF predictions, I sought to understand whether there were any links with myelopoietic stimuli. I examined the plasma levels of myelopoietic cytokines in cohort 2 samples and found that M-CSF and GM-CSF were not differential across SRS, whereas G-CSF and IL-6 which canonically signal through STAT-3 and are known to be key contributors to EG (Zhang et al. 2010; Walker et al. 2008) were elevated in SRS1 (Fig. 5.32). This pattern of myelopoietic cytokines matched the observed peripheral blood cellular compositional changes of increased immature neutrophils but no increase or even a decrease in monocytes in SRS1 (Fig. 5.14).



**Figure 5.32: Myelopoietic cytokines in sepsis.** Plasma myelopoietic cytokine levels by SRS from cohort 2 samples (SRS1 n=22, non-SRS1 n=14) (two-sided Wilcoxon rank-sum test).

Previous reports linked G-CSF priming of neutrophils in cancer to increased neutrophil NETosis (Demers et al. 2012). On measuring sepsis neutrophil NETosis responsiveness to stimulus (Phorbol 12-myristate 13-acetate, PMA), I found that SRS1 neutrophils underwent more NETosis than non-SRS1 neutrophils (Fig. 5.33), concordant with the plasma cytokine data.



**Figure 5.33: NETosis in sepsis neutrophils** (A) Representative images of neutrophils undergoing NETosis from SRS1 and non-SRS1 patients (four hours post-PMA stimulation, green fluorescence denotes cells undergoing NETosis). (B) NETosis in neutrophils across SRS (n=10 in each group, two-sided Wilcoxon rank-sum test). FDR = false discovery rate.

## 5.3 Discussion

The experiments presented in this chapter demonstrate the feasibility of interrogating neutrophils with single cell sequencing technologies despite the logistical challenges associated with fresh sample processing. The data presented reinforce the specificity of immature neutrophils to sepsis, demonstrate the immunosuppressive nature of neutrophils in general in sepsis, and show that SRS1 represents a state of particularly heightened immature neutrophilia and more substantial neutrophil immunosuppression. The elucidation of the underlying cellular immunological and functional differences across SRS states consolidates the endotype definition ascribed to these transcriptomic subgroups.

### 5.3.1 Neutrophils/granulocytes in sepsis

The current data represent one of the most comprehensive catalogues of neutrophil phenotype and function in sepsis and in particular in resolving sepsis heterogeneity. Neutrophil single cell sequencing datasets thus far have been smaller in scale, both in sample and cell numbers, and mostly in different contexts of health and disease (Schulte-Schrepping et al. 2020; Sinha et al. 2021). Moreover, previous neutrophil single cell sequencing studies have not validated the functional aspect of identified immunosuppressive phenotypes while the data presented confirm that neutrophils in sepsis are indeed functionally immunosuppressive.

Most studies of MDSCs have come in the context of cancer biology where there is interest in harnessing the immune system to destroy tumors, or in the case of infectious diseases often in animal models (Tebartz et al. 2015; Dietlin et al. 2007). Indeed, there has been significant progress in the form of T cell based immunotherapies for cancer treatment in the recent decade (Waldman et al. 2020), but corresponding strategies to leverage MDSC knowledge has been lacking, probably owing to the much shallower understanding of MDSCs as a whole.

Here, I provide evidence for the relevance of G-MDSCs in a context beyond cancer and animal models of infectious disease i.e. in sepsis samples from humans. However, while I identify shifts in neutrophil distributions, phenotype and function in sepsis and across SRS, it is nevertheless notable that the neutrophils interrogated are from the peripheral blood rather than the site of infection source. This is important because the immune response shows properties of compartmentalization. For example, neutrophils in mouse models of LPS administration were activated (with increased pro-inflammatory cytokine production) in the lungs, but not in the blood (Abraham et al. 2001)). Similarly,

in a mouse model of peritonitis, PBMC TNF production was reduced, but tissue resident macrophages in the lungs (alveolar macrophages) and the liver (Kupffer cells) had enhanced TNF production (Suzuki et al. 2006). These data highlight the importance of considering the function of the cells at the site of pathology.

A previous scRNA-seq study on sepsis PBMCs highlighted an immunosuppressive monocyte state, MS1, associated with increased mortality and likely derived from altered myelopoiesis (Reyes et al. 2020; Reyes et al. 2021). The data from this chapter however suggest that the underlying contribution of circulating monocytes to the sepsis endotypes as defined by whole blood transcriptomics is relatively minor across all three cohorts. Given the similarity of genes defining the MS1 state and our various neutrophil populations, it is tempting to speculate that deconvolution analysis by the original authors using a PBMC scRNA-seq reference on whole blood data would have mistaken whole blood neutrophil populations for monocytes in the reference.

### **5.3.2 Mechanism of G-MDSC immunosuppression**

The current data suggest that the commonly cited mechanisms for MDSC immunosuppression through the PD-1 - PD-L1/L2 pathway or through MDSC arginase-1 depletion of arginine for T cells are not the immunosuppressive mechanisms at play in sepsis. While inhibition of the PGE2 receptor achieved partial restoration of T cell activation, our experiment was limited by a relatively small sample size, and the inability to definitively pinpoint the relevance of PGE2 given that inhibition of cyclo-oxygenase (COX) enzymes which produce prostaglandins did not achieve the same partial reversal. While it is possible that it is indeed PGE2 that is involved and that the mechanism of T cell immunosuppression is through pre-formed PGE2, hence accounting for the failure of COX inhibition, more evidence is still needed before this hypothesis can be confirmed e.g. staining for PGE2 within sepsis neutrophils at the beginning

of and after co-culture with T cells.

Alternatively, the G-MDSC immunosuppression may be mediated through other mechanisms not tested in the experiments above e.g. immunosuppressive cytokines such as IL-10 and TGF- $\beta$ , or nitric oxide (Umansky et al. 2016; Veglia et al. 2021). Another potential mechanism is the depletion of tryptophan by indoleamine-2,3-dioxygenase (IDO) by MDSCs (Yu et al. 2013; Köstlin et al. 2017, Platten et al. 2019); tryptophan is an essential metabolite for T cell function. Modulating these pathways in further sepsis/SRS neutrophil-CD4<sup>+</sup> T cell co-cultures would help elucidate the exact mechanism of T cell suppression observed.

### **5.3.3 Ontogeny of suppressive/immature neutrophils**

The expansion of immature neutrophils in sepsis and in particular in SRS1, also brings the question of the origin and ontogeny of these cells. As previously mentioned, MDSCs are often associated with immaturity, but the current data suggests that both mature and immature neutrophils in sepsis and SRS1 are immunosuppressive.

In the healthy or steady-state, granulopoiesis is mainly governed by C/EBP $\alpha$  (Manz and Boettcher 2014). However, in settings of physiological stress such as trauma or infection, a process known as emergency granulopoiesis (EG) is activated to increase neutrophil production to meet the demands for neutrophils in the periphery upon neutrophil consumption. Granulopoiesis under these conditions is mainly enabled by C/EBP $\beta$  (Hirai et al. 2006). Upstream of this, pathogen sensing is thought to primarily be mediated by non hematopoietic stem cell (HSC) pattern recognition receptor (PRR) signaling even though HSCs themselves express PRRs (Nagai et al. 2006; Schmid et al. 2011). The *in vivo* physiological role of HSC PRR ligation is not fully understood; this was demonstrated through extensive knockout mouse experiments where tissue

specific MyD88 deletion was performed for multiple compartments, leading to the conclusion that endothelial cells were the particularly important cell type for pathogen sensing (Boettcher et al. 2014).

After pathogen detection, the molecular events that lead to EG include G-CSF and IL-6 signaling through STAT3 (Zhang et al. 2010; Walker et al. 2008), the latter cytokine identified through the observation that neutrophil promoting activity exists in media conditioned from cells of G-CSF and GM-CSF double knockouts. In addition to C/EBP $\beta$  and STAT3,  $\beta$ -catenin-T cell factor / lymphoid enhancer-binding factor (TCF/LEF) and the Wnt signaling pathway were also shown to be important for EG by upregulating the G-CSF receptor (Danek et al. 2020).

Interestingly, a series of reports of MDSCs in a sepsis mouse model have identified an important molecular circuit downstream of C/EBP $\beta$  related to expansion of sepsis MDSCs (Brudecki et al. 2012; McClure et al. 2014; McClure et al. 2016; McClure et al. 2017; McPeak et al. 2017). Using a polymicrobial CLP sepsis model, Brudecki et al (2012) demonstrated peripheral blood MDSC expansion as sepsis progressed, with nuclear morphological appearances and cell surface CD31 expression supporting immaturity in the MDSCs identified (Ling et al. 1997). Subsequently, it was demonstrated that the microRNAs miR-21 and miR-181b were responsible for expanding these MDSCs (McClure et al. 2014), and that the transcription factor NFI-A blocked maturation of the MDSCs (McClure et al. 2016). Next, it was shown that the TFs Stat3 and C/EBP $\beta$  induced expression of the aforementioned microRNAs (McClure et al. 2017), and that C/EBP $\beta$  knockout mice did not produce MDSCs.

In the context of the current data, the prediction of STAT3 as a regulator for multiple gene programs more highly expressed in SRS1 in various neutrophil states in conjunction with the elevated levels of plasma IL-6 and G-CSF suggest that SRS1 is a state of heightened EG. This is also consistent with the observation

of C/EBP $\beta$ , the master regulator of EG, also being a predicted TF regulator for the same gene programs that show STAT3 target enrichment. To interrogate the hypothesis of SRS1 (or high SRSq scores) being a state of EG, the next step should involve experiments designed to profile the hematopoietic stem and progenitor cell (HSPC) compartments in patients with SRS status assigned. The more definitive data would ideally be derived from bone marrow samples, but given that bone marrow sampling in sepsis patients is not performed as part of routine sepsis clinical care, an alternative could be to profile circulating HSPCs. This would undoubtedly require substantial enrichment of these cells before profiling as HSPCs represent only ~0.1-2% of circulating PBMCs.

### **5.3.4 Therapeutic opportunities**

The data from this chapter raise a number of potential therapeutic targets for sepsis. First, there is a clear opportunity for G-CSF and/or IL-6 modulation in SRS biomarker-led sepsis clinical trials. Previous clinical trials of G-CSF in sepsis failed to reduce mortality (Root et al. 2003; Bo et al. 2011; Stephens et al. 2008), while trials of GM-CSF are ongoing (Pinder et al. 2018). However, an improved understanding of the mechanisms by which these drugs influence neutrophils and granulopoiesis, and which patients should be targeted, will be key to success. The addition of granulopoietic stimuli may not enhance the functional capacity of circulating immature and immunosuppressive cells, and the effects on what is possibly heightened EG in SRS1 is unknown. In view of increased levels of G-CSF and IL-6 and immunosuppressive neutrophils in SRS1, inhibition of these two cytokines in SRS1 patients represents a potential immunotherapeutic strategy to boost immune function. The echoes of the clinically positive result for IL-6 inhibition in a specific group of COVID-19 patients i.e. those with hypoxia (oxygen saturation <92% on room air or requiring oxygen therapy) and signs of systemic inflammation (C-reactive protein  $\geq 75$  mg/L) (Abani et al. 2021) are encouraging in terms of reinforcing the importance of correct selection of

patients, and the parallel of IL-6 inhibition in these patients makes the prospect of a similar strategy in sepsis an attractive target.

Similarly, the data above suggest that modulation of the IL-1 pathway may be worth revisiting in biomarker-led sepsis clinical trials. While the plasma cytokine levels of either IL-1 $\alpha$  or IL-1 $\beta$  were not differentially elevated in SRS1, the mature neutrophil gene expression program enrichment in response to IL-1 and expansion of *IL1R2*<sup>+</sup> immature neutrophils in SRS1 suggest an enhanced reaction to IL-1 even if absolute levels are not increased. In addition, it is not possible to understand from the peripheral blood and plasma data what local IL-1 concentrations may be either at the site of the infection or at the bone marrow where EG may be enhanced in SRS1, especially given the known phenomenon of IL-1 $\alpha$  being mostly associated with the plasma membrane (Kurt-Jones et al. 1985, Sims and Smith 2010), leaving open the possibility that selective targeting of IL-1 in a spatiotemporally appropriate manner may be beneficial.

While trials of IL-1 inhibition have previously failed to demonstrate mortality reduction (Opal et al. 1997; Fisher et al. 1994), these trials suffered from the absence of patient stratification. In a re-analysis of a phase III trial of IL-1R blockade with anakinra, improved survival was seen in a subgroup of anakinra-treated patients with hepatobiliary dysfunction and disseminated intravascular coagulation (Shakoory et al. 2016), suggesting the need for identification of the correct target patient group who may benefit from the therapy. The current data uncovering the basis of SRS endotypes suggest that there is potential utility in prospective trials of IL-1 inhibition in SRS1 patients specifically. However, unlike IL-6 inhibition where a success story has already come in the context of severe COVID-19, IL-1 inhibition in COVID-19 has not seen similar efficacy (Derde et al. 2021; Caricchio et al. 2021; Mariette et al. 2021) despite originally promising observational data (Aouba et al. 2020; Cavalli et al. 2020) and early phase clinical trial data (Huet et al. 2020; Kooistra et al. 2020), making IL-6/G-



CSF higher priority targets than IL-1 modulation.

## 5.4 Conclusions

This chapter has taken previous whole blood defined bulk transcriptomic signatures and dissected the cellular contributions, gene expression programs and likely mediators that underlie the signatures. This has refined our biological understanding of the described sepsis endotypes, highlighting targets for future experimental manipulation to definitively establish causal mechanisms, as well as nominating potential therapeutic targets for precision medicine in sepsis.

## GENERAL DISCUSSION

*This chapter outlines the broader conclusions and future directions suggested by the work described in this thesis.*

|     |                                                                          |     |
|-----|--------------------------------------------------------------------------|-----|
| 6.1 | Multi-omics . . . . .                                                    | 207 |
| 6.2 | Tissues of interest in severe infection . . . . .                        | 212 |
| 6.3 | Tissues of interest in the context of legacy effects of severe infection | 214 |
| 6.4 | Animal models for SRS . . . . .                                          | 216 |
| 6.5 | Translational opportunities . . . . .                                    | 217 |
| 6.6 | Limitations and future work . . . . .                                    | 219 |
| 6.7 | Conclusion . . . . .                                                     | 222 |

This thesis has applied an immune multi-omic approach to tackle the heterogeneity underlying severe infection and sepsis across a range of pathogen contexts. Specifically, I have extended previous peripheral blood bulk transcriptomic definitions (sepsis response signatures, SRS) to new cohorts of patients including novel causes of sepsis e.g. SARS-CoV-2 induced viral sepsis in COVID-19 (chapter 3), and dissected underlying immune mechanisms through single cell multi-omics in both the peripheral blood mononuclear cell (PBMC) compartment (chapter 4), as well as in the total leukocyte population of peripheral blood (chapter 5), elucidating the drivers of SRS.

The implications of the results are considered below in the context of current understanding of sepsis, with potential areas for follow up work discussed.

## 6.1 Multi-omics

### Atlasing strategies and further multi-omics

As human biology continues to move towards the construction of cell atlases for health and disease, datasets such as those presented in chapters 4 and 5 will become invaluable resources, including in the context of unravelling cellular specificities that contribute to bulk sequencing datasets such as that presented in chapter 3. Cell atlases based on single cell sequencing technologies have been built for other diseases, such as systemic lupus erythematosus, colorectal cancer and Alzheimer's disease. These atlases can be leveraged for future studies with cell identities mapped back on to current datasets to improve consistency across experiments in terms of cellular definitions.

The key differences between the two single cell atlas-level dataset resources presented are the more extensive comparator groups included in the PBMC dataset, and the inclusion of granulocytes in the whole blood dataset. The PBMC dataset fused extensive pre-existing knowledge regarding protein markers of immune subsets with the rich information from a data driven approach in the gene expression space, revealing heterogeneity in cell states that was previously underappreciated. Future atlases will undoubtedly frequently employ a similar multi-modal profiling platform to elicit both RNA and protein data simultaneously. The whole blood dataset brings significant value in understanding neutrophil biology, as neutrophils are a less studied immune cell type, and the difficulty of scRNA-seq in neutrophils means that few datasets interrogating neutrophils exist, especially in the disease context.

More generally, the feasibility of large scale single cell datasets as presented in this thesis highlight the opportunity for future single cell atlasing approaches, both in terms of scale, and in terms of modalities. Single cell profiling technologies now extend beyond RNA and protein to chromatin, histone mark

and DNA methylation data, and the combination of multiple modalities (Ma et al. 2020; Clark et al. 2018).

### **Multi-omics integration**

The data from the thesis also bring up the key question of multi-omic data integration. The thesis has presented gene expression and protein abundance data in various forms and integrated them with different methods including sequential analysis in chapter 4, and variational autoencoder frameworks and factor analysis in chapter 5.

The factor analysis approach employed in integrating bulk RNA-seq and mass cytometry (CyTOF) cell proportion data represented an example of the power of linear algebraic methods. The factor analysis implementation used (Multi-Omic Factor Analysis plus, MOFA+) has been described as a multi-view or multi-dimensional generalization of principal components analysis (Argelaguet et al. 2020). Similar approaches exist such as Sparse Decomposition of Arrays (SDA) (Hore et al. 2016), Integrative Non-negative Matrix Factorization (intNMF) (Chalise and Fridley 2017), Sparse multiple Canonical Correlation Analysis (SmCCA) (Witten and Tibshirani 2009), and Joint and Individual Variation Explained (JIVE) (O’Connell and Lock 2016), which are all fundamentally variations of linear algebraic matrix or tensor decomposition methods with a variety of extra constraints or statistical features built into the models to suit biological multi-omic data, each with various advantages and disadvantages.

For example, while CCA only allows for common sources of variation across data types to be inferred, the other methods permit both individual and joint sources of variation to be inferred. While both SDA and MOFA+ implement sparsity priors in order to best identify key contributors to the factors or components decomposed, the NMF technique has a constraint that factorized units must be non-negative, which suits data types such as gene expression data where a

gene can be repressed at most to 0 expression, but not a negative expression value. The models in JIVE, MOFA+ and SDA can all cope with missing values for certain samples, but MOFA+ has the greatest flexibility in terms of support for a variety of likelihoods to suit the different distributions that different -omic data types provide, including Gaussian, Poisson and Bernoulli likelihoods which for example would be suitable for gene expression, count data and mutation data respectively. Table 6.1 breaks down the properties of the different implementations .

| Technique | Underlying approach               | Sparsity assumptions | Model distributions | Bayesian     | Non-negativity | Software implementation |
|-----------|-----------------------------------|----------------------|---------------------|--------------|----------------|-------------------------|
| MOFA+     | Factor analysis                   | Yes                  | Flexible            | Bayesian     | No             | R package MOFA+         |
| SDA       | Tensor decomposition              | Yes                  | Fixed               | Bayesian     | No             | SDA software            |
| intNMF    | Non-negative matrix factorization | No                   | Unspecified         | non-Bayesian | Yes            | R package intNMF        |
| SmCCA     | Canonical correlation analysis    | Yes                  | Unspecified         | non-Bayesian | No             | R package PMA           |
| JIVE      | Principal component analysis      | No                   | Unspecified         | non-Bayesian | No             | R package r.jive        |

**Table 6.1: Multi-omic integration methods.**

A key advantage of linear methods is interpretability. However, relationships in biology may not necessarily be linear, and use of non-linear methods for multi-omic integration in the context of single cell sequencing has been particularly extensive. In chapter 4 of the thesis, a sequential approach of first splitting cell types based on protein marker defined lineages before clustering on gene expression was utilized to ensure purity of lineage definitions. However, there is also increasing evidence that in certain immune cells e.g. T cells, definitions based on canonically well understood markers such as CD4 and CD8 may artificially discretize cell states that are possibly more continuous in nature (Nathan et al. 2022).

To best understand potential non-linear aspects to the data while avoiding issues of cell type/state discretization due to over-weighting of protein markers based on prior knowledge, there have been a number of single cell multi-omic methods developed based on autoencoder frameworks (Lopez et al. 2018; Amodio et al. 2019; Tran et al. 2021). Chief amongst these has been a suite of software known as scvi-tools (Gayoso et al. 2022), which began with scRNA-seq analysis (Lopez et al. 2018) and has extended to incorporate multi-omic analysis at RNA and protein levels (Gayoso et al. 2021) (used for the single cell dataset in chapter 5) as well as RNA and chromatin levels (Ashuach et al. 2022).

The basic model employed within these methods is based on a variational autoencoder (VAE) framework where raw data is fed into VAEs parametrized with multiple neural networks configured to explicitly account for known technical covariates e.g. batch, as well as experimental properties e.g. zero-inflated negative binomial distributions for scRNA-seq data, and Gaussian mixture models for cell surface antibody profiling to account for foreground antibody staining (true protein expression) and background antibody staining effects (batch and cell specific). The use of neural networks to learn latent representations for each cell which should approximate the true biological

manifold the cells lie in allows for non-linear properties to be uncovered, and the simultaneous consideration of multiple data modalities for learning a single set of latent representations per cell achieves multi-modal integration.

The two single cell datasets generated in this thesis present opportunities for all the methods mentioned above to be benchmarked against each other as well as for further method development, for example to integrate protein, RNA and chromatin data all together simultaneously. This could be done with either newer VAE methods which account for all three layers of data, or linear algebraic methods such as MOFA+. Reanalysis of the datasets with some of the methods mentioned above represents an opportunity for further biology to be uncovered beyond what has been presented.

## **6.2 Tissues of interest in severe infection**

This thesis has focused on using blood as a window to the study of the immune response to severe infection and in sepsis. While useful for logistical reasons in terms of ease of obtaining patient samples and relevant given that immune cells are present in blood, this nevertheless misses the action either at sites of infection source and pathology, or other tissue sites potentially relevant to disease. Broadly speaking, this is important because of the compartmentalization of immune reactions (Cavaillon and Annane 2006). One of the classic examples of this has been the observation that while blood monocytes undergo endotoxin tolerance in either mouse models or humans from lipopolysaccharide (LPS) administration, alveolar macrophages are instead primed, which was shown in mouse to be due to the lung microenvironment and cytokine milieu from B cells (Philippart et al. 2012), with the priming phenomenon and also seen in humans (Smith et al. 1994; Hoogerwerf et al. 2010).

An example of a relevant tissue of interest would be lung tissue in the context



of COVID-19, influenza or more generally community acquired pneumonia (CAP) sepsis. Indeed there have since the time of data generation of the current PBMC atlas a number of scRNA-seq datasets of respiratory samples of COVID-19 patients, permitting a tissue specific view of the disease. Liao et al (2020) provided the first such report, sequencing immune cells from bronchoalveolar lavage fluid (BALF) samples, finding expansion of both proinflammatory macrophages and CD8<sup>+</sup> T cells compared to healthy donors. The latter result mirrored the expansion of effector CD8<sup>+</sup> T cells seen in blood in COVID-19 patients, but whereas the blood samples also showed an expansion of effector CD4<sup>+</sup> T cells, the dataset from Liao et al did not. The BALF dataset also highlighted the relevance of macrophages, which do not exist in the blood, with the authors suggesting that macrophages in severe disease may contribute to further inflammation in critical COVID-19 through CCR1 and CXCR2 mediated recruitment of monocytes and macrophages. However, the origins of these macrophages studied in terms of whether they were tissue resident alveolar macrophages vs. monocyte derived macrophages was not determined. The dataset also had a very small sample size, containing only three patients with moderate COVID-19 (defined as patients with bilateral pneumonia), one with severe disease (defined as patients requiring oxygen support) and five with critical disease (defined as patients requiring mechanical ventilation).

In a different study setup but also interrogating respiratory samples, Chua et al (2020) performed scRNA-seq on nasopharyngeal (upper respiratory tract) and bronchial (lower respiratory tract) samples from 19 COVID-19 patients and 5 healthy controls. Similar to the BALF dataset, they again found more inflammatory macrophage and CD8<sup>+</sup> T cell activity in COVID-19 patients with more severe disease. This was seen as a stronger interaction in these patients between epithelial cells and immune cells with the macrophages expressing a range of chemokines and cytokines including classically pro-inflammatory IL-1 $\beta$

and TNF, as well as CCL2, CCL3, CCL20, CXCL1, CXCL3, CXCL10 and IL-8. These immune cell changes were seen in both the upper and lower respiratory tract samples.

However, neither of these respiratory sample datasets included paired blood samples, precluding direct conclusions of differences in the two compartments. Future paired -omic datasets will therefore be of immense value for better understanding the anatomical specializations exhibited by the immune cells.

Another example of a set of tissues that would be worth profiling in the context of severe infection would include the secondary lymphoid organs. Lymph nodes (LNs) are sites of antigen presentation, lymphocyte clonal selection and expansion and germinal centre reactions for developing humoral immune responses to infections. Obtaining either cellular aspirates of LNs for immunophenotyping or single cell sequencing, or biopsies for histology and potentially spatial transcriptomics would be highly valuable samples for better understanding the dynamics of the development of an adaptive immune response during sepsis. While this has been done in the context of other infectious diseases e.g. tuberculosis (Maji et al. 2015), there have been no such studies in humans in sepsis or COVID-19 and is a potential future avenue of investigation.

### **6.3 Tissues of interest in the context of legacy effects of severe infection**

One aspect of biology in severe infection not considered in this thesis is the legacy of the infection. In sepsis it is well known that a substantial proportion of survivors suffer from post-sepsis syndrome and that survivors in general have higher morbidity and mortality in both short and medium-long term compared to age matched controls. COVID-19 meanwhile has been widely reported to lead

to a variety of debilitating symptoms after the acute course of illness, known popularly in the media as "long COVID" and medically as post acute sequelae of COVID-19 (PASC).

However, in studies with extensive symptomalogical assessment of PASC, there were no clear risk factors identified for PASC by either clinical blood tests or peripheral blood immunological evaluation in terms of antibody, T cell or cytokine levels (Sneller et al. 2022). While the cause of PASC is unclear and may be due to viral-independent mechanisms (Nalbandian et al. 2021), there has also been growing evidence of potential viral persistence with various tissues potentially being relevant. First, in the context of LNs, the phenomenon of antigen archiving (Kedl and Tamburini 2015) whereby lymphatic endothelial cells may serve as reservoirs of viral antigen long after viral clearance is a plausible mechanism for retention of SARS-CoV-2 antigen after infection. While there has been no direct interrogation of this in PASC, it has been observed that SARS-CoV-2 spike protein can be found in axillary LNs up to 60 days after the second dose of an mRNA COVID-19 vaccine (Röltgen et al. 2022).

Another potentially relevant tissue is the gut. It was initially demonstrated that in patients with COVID-19 diagnosed four months prior, there was continuing evidence of B cell antibody evolution, suggesting the continued presence of viral proteins (Gaebler et al. 2021). In a subsequent study of patients with inflammatory bowel disease, it was then shown that there was presence of viral antigen in the gut of these patients seven months post infection (Zollner et al. 2022), backing up previous data of fecal COVID-19 RNA shedding long after acute COVID-19 infection (Natarajan et al. 2022). More generally, there has been preliminary evidence for the persistent presence of viral RNA in other anatomical locations including the heart, eyes and brain from autopsy studies (Chertow et al. 2021).

Finally, given the findings of immature neutrophilia in the clinically deleterious

SRS1 state and the increase in plasma levels of emergency granulopoiesis (EG) cytokines in SRS1, an obvious candidate tissue for further study in sepsis is the bone marrow. The bone marrow is a clear candidate for future investigation into contributions of EG in sepsis and post sepsis given that the lifespan of mature myeloid cells is only at most up to one week, and both phenomena of endotoxin tolerance and trained immunity occur via epigenetic priming of bone marrow progenitor cells. Indeed, there have been recent investigations into the links between bone marrow effects and PASC, with signatures of AP-1 and interferon pathways deemed relevant.

## **6.4 Animal models for SRS**

One key bottleneck in studying the above tissues of interest is the difficulty in obtaining such samples from human patients. There is therefore substantial value in considering relevant animal models to build on and complement the work presented in this thesis.

In the context of SRS, it will be important to develop a relevant model that can faithfully recapitulate SRS endotypes, and allow experimental manipulation to establish causality and likely directions of effects on either inhibition or boosting of specific pathways. From the data presented in chapter 5, it is clear that SRS definitions are heavily neutrophil based. This will create difficulty in using mouse models as only roughly 25% of circulating leukocytes in mice are neutrophils while the other 75% are MNCs i.e. the exact reverse of humans. However, the fact that the murine immune system is well studied and understood makes it an attractive model to be used, especially as sepsis models are well established already such as the cecal ligation and puncture (CLP) model. It will be important to ensure that the model is appropriate to the SRS patients studied, e.g. the use of aged mice as well as the provision of a certain degree of resuscitation after onset of sepsis would provide a much more realistic parallel.

An alternative model which would be viable for studying sepsis is the *Danio rerio* model organism i.e. the zebrafish. The zebrafish represents an accurate model of human hematopoiesis (Trede et al. 2004, and models have been used for high-throughput drug discovery for resistance to LPS administration (Philip et al. 2017). The size of the zebrafish also permits evaluation of not just the immune system but effects on the whole organism. Specifically in the context of EG, zebrafish models have been used to elucidate potential mechanisms relevant to the process, such as the importance of macrophage derived IL-1 $\beta$  signaling through CEBP $\beta$  for EG (Wei et al. 2020) and the role of inducible nitric oxide synthase downstream of CEBP $\beta$  in expansion of hematopoietic stem and progenitor cells (Hall et al. 2012).

## **6.5 Translational opportunities**

The extensive -omic profiling in datasets such as those presented in this thesis present a variety of translational opportunities, from diagnostic/prognostic stratification to therapeutic ones. Below is a summary of themes that have emerged from the thesis beyond those already discussed in individual chapters.

### **6.5.1 Predicting SRS with clinical parameters**

Given the robustness of SRS in stratifying patients for early mortality and its applicability to infections of different organisms, there is value in utilizing this definition by the bedside. While the current definition using seven genes should be amenable for clinical deployment e.g. through a rapid quantitative polymerase chain reaction (qPCR) platform, there would be even greater feasibility if SRS could be derived from clinical parameters.

One approach to this problem would be to obtain extensive clinical data, both clinical measurements such as symptoms and bedside observations as well as blood test results, and perform feature selection e.g. through a sparse lasso

or elastic net regression model. The advantage of this would be its relatively unbiased and data driven nature, while the drawback may be that selected parameters are not easily interpretable.

Alternatively, a biologically motivated method of feature selection could be used, for example, by considering the results from chapter 5. The elevation of IL-6 in SRS1 suggests that SRS is to some extent governed by the acute phase response. The immature neutrophilia in SRS1 suggests a differential bone marrow response by SRS to the EG cytokines of IL-6 and/or G-CSF. Together, there is therefore rationale for selecting markers which may reflect both of these aspects. The first could be considered through positive and negative acute phase reactants (C-reactive protein and albumin respectively), while the second could be measured by the neutrophil-lymphocyte ratio from peripheral blood count tests. Both sets of markers are routinely and frequently measured in hospitalized and especially sepsis patients, and if achieving sufficient accuracy in SRS assignment, would be easily and immediately usable clinically.

### **6.5.2 Therapeutic opportunities**

In addition to SRS status assignment for prognostication, there are a number of therapeutic implications to the datasets presented in the thesis. As discussed in chapter 5, G-CSF, IL-6 and IL-1 represent potential targets for clinical trials with prospective patient stratification, although animal modeling as discussed above would be key to ensure that the correct direction of effect is established before moving on to clinical trials.

The use of steroids is also worth considering in the broader context of SRS and severe infections. As was demonstrated in the RECOVERY and other trials, steroids are clearly useful for patients with severe COVID-19. However, their use in sepsis is much more controversial with conflicting data in terms of outcomes. This discrepancy may reflect an issue of tissue specific effects,

where in COVID-19 the main site of pathology is in the lungs and dampening of hyperinflammation in the lungs may be beneficial, whereas when considering sepsis as a whole, there is a lack of such a specific target or site of potential action.

While the post hoc analysis of VANISH showed an interaction effect between steroid administration and SRS (Antcliffe et al. 2019), it suggests that steroids may in fact be harmful in non-SRS1 patients, but did not show benefit in any groups. In the context of the immature neutrophilia seen in SRS1, it is worth considering that steroids are known to increase immature neutrophils in the peripheral blood, both by demargination of existing neutrophils in circulation, but also by stimulating granulopoiesis. It is possible therefore that steroids may be mediating a harmful effect in non-SRS1 patients if their effect is to produce an immature neutrophilia similar to that which defines the difference between SRS1 and non-SRS1 patients. Notwithstanding this, the VANISH result remains to be replicated and validated in larger and prospectively cohorts.

Finally, aside from steroids and cytokine modulation, the work in this thesis suggest other candidates for therapeutic intervention in severe infection. Metabolism has been reported as a key aspect of SRS differences (Davenport et al. 2016), and scRNA-seq data from chapter 5 show that the relevant cell type for this are *IL1R2*<sup>+</sup> immature neutrophils which display heightened glycolysis. This could be modulated with glycolysis inhibitors, but this would require functional experiments first to demonstrate potential efficacy before moving on to clinical trial settings, and the ability to specifically modulate glycolytic activity within the right cell type would need to be established.

## 6.6 Limitations and future work

There are necessarily an array of limitations to the data and work presented in this thesis. While certain limitations have already been covered, the broader

limitations are discussed below.

First, it is worth reiterating that in observational data as presented, it is not possible to delineate which of the changes seen are causal for vs. a consequence of disease. One improvement would be to perform a time course analysis with dense sampling e.g. every other day. Alternatively, the development and maturation of the field of causal inference as applied to observational data may also mitigate this issue. However, even such time course data or novel analytical techniques would not fully resolve causality given potential for unobserved confounding. Perturbation remains the gold standard for establishing causal mechanisms, leaving a fundamental need for experiments. The proposal of potential animal models above is greatly motivated by this, but a direct opportunity for this in humans exists in terms of collecting samples from clinical trials for analysis. At time of writing, there are 250 actively recruiting trials for sepsis registered on [www.clinicaltrials.gov](http://www.clinicaltrials.gov), but the vast majority do not stipulate collection of samples for deep molecular -omic profiling within their protocols. The infrastructure and funding for the setup as was the case in the VANISH trials could therefore be applied to these ongoing clinical trials and would uncover a wealth of information regarding causal mechanisms in sepsis in humans.

Second, as discussed above, one of the major limitations in the data in this thesis is that only blood samples were used. There is therefore an imperative to ensure that other potentially relevant tissues are sampled and profiled, with the hematopoietic compartments one of the key priorities.

Third, this thesis has heavily deployed a transcriptomic lens with aspects of protein level data also collected and analyzed, but other axes of interrogation are also needed. Given the poor correlation between transcript and protein levels and that proteins mediate the majority of biological function, future proteomic assays would be able to add substantial value beyond the current datasets. Furthermore, proteins are subject to extensive regulation in terms of either



degradation or also post-translational modifications, for which additional data will be needed to achieve a complete description of proteomic contributions in severe infection. The regulation of gene expression is also important to consider, for example, through epigenetic assays such as chromatin and histone mark profiling with Assays for Transposase Accessible Chromatin sequencing (ATAC-seq) and Chromatin Immunoprecipitation sequencing (ChIP-seq) respectively. In the original description of SRS, expression quantitative trait loci (eQTL) mapping was performed. In that analysis, there was no information on cellular specificity, and as such it could not be ascertained to what extent the identified relationships between eQTLs and SRS were due to cell type/state proportion changes. In the current analysis, the sample sizes for either eQTL analysis in bulk RNA-seq data in COVID-19 or in the scRNA-seq datasets for COVID-19 and non-COVID-19 sepsis were too small. Future studies of larger sample sizes for such eQTL mapping are therefore a potential avenue of future work.

Even within transcriptomics, there are aspects that have yet to be examined. Messenger RNA (mRNA) molecules are processed after transcription, with splicing being one of the key biological processes that governs a diversity of downstream function. While splicing information was considered in the context of trajectory analysis by RNA velocity in assessment of neutrophil maturity, explicit testing of differential splicing across the various different conditions of severe infection was not done for this thesis. Transcript stability is also subject to regulation which was also not assessed and could be done e.g. through chromatin-associated RNA enrichment (Wissink et al. 2019). Moreover, this thesis has focused mostly on mRNA analysis, but other RNA molecules that are part of the transcriptome such as micro RNAs and long non-coding RNAs could be profiled in future studies.

Fourth, in considering sepsis in general, it will be important to put more focus on microbiological causes and the host-pathogen interaction. While the thesis

has demonstrated the utility of SRS in a broad range of infectious aetiologies, it is nevertheless the case that microbial etiology does affect the host response (Sweeney et al. 2016; Burnham et al. 2017; Goh and Knight 2017). Major challenges for this include the fact that 50% of sepsis cases are culture negative, and that a wide range of organisms can cause sepsis, meaning it will be difficult to achieve adequately sized and therefore powered studies for stratification by microorganism. Recent developments in improving pathogen detection in infections by targeted metagenomic sequencing are a potential solution to this issue (Goh et al. 2019).

Finally, it is worth considering that despite the various potential therapeutic avenues that could be explored given the data from the thesis, the attrition rate from the current stage to success at the bedside is astronomical. Within clinical trials alone, the estimated success rate from phase I clinical trials to approval of a drug and marketing is under 20%. The rate of success from the current exploratory data which will require in vivo validation before moving to phase I clinical trials is harder to measure but likely to be even lower. Together, this is a reminder of the base rate of success in moving the promising targets from this thesis to eventual clinical application.

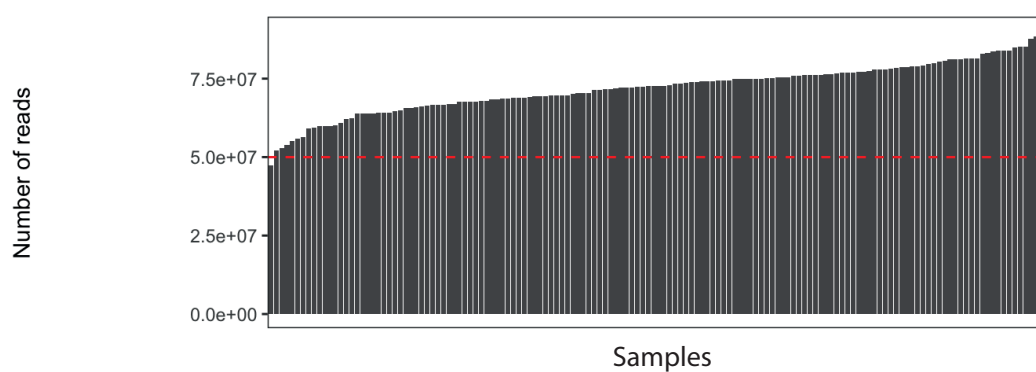
## **6.7 Conclusion**

Notwithstanding the discussed limitations, there are a variety of ways which work from the thesis has advanced understanding of the immune response to severe infection. I have produced a number of resources of sepsis -omic data and set up a new study and associated sample bank for future experiments to build on the current results. The current body of work has broadened the application of transcriptomic endotype definitions and examined specific cellular immunological changes and contributions to the response to severe infection. These findings highlight opportunities for precision medicine in sepsis with

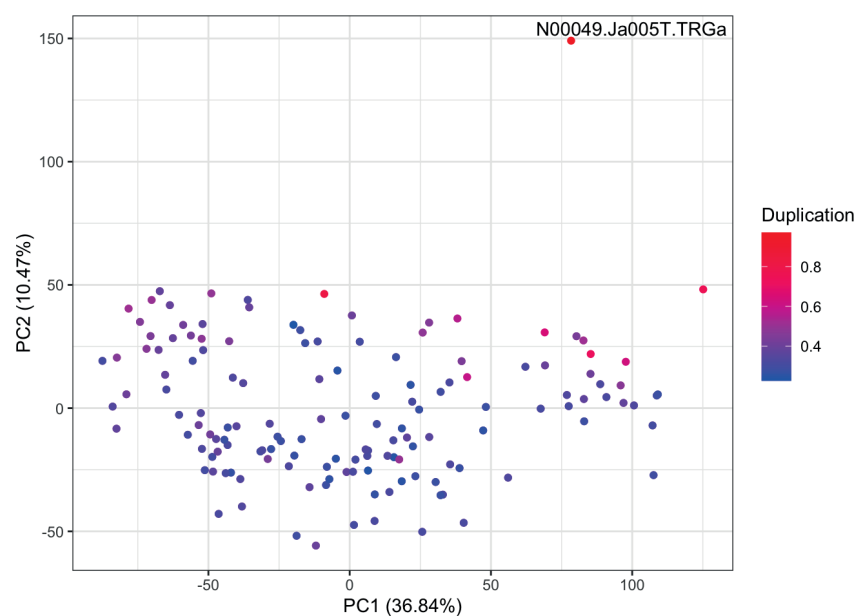
nomination of potential therapeutic targets.

## APPENDIX TO CHAPTER 2

A



B



**Figure A.1: Bulk RNA-seq quality control (QC).** (A) Number of sequencing reads per sample. Red dotted line shows target of 50 million reads. (B) Principal component analysis of  $n=144$  samples with one outlier sample on PC2 shown. Samples colored by read duplication rate.

| Surface protein                | Clone     |
|--------------------------------|-----------|
| anti-Tau Phospho (Thr181)      | M7004D06  |
| B7-H4                          | MIH43     |
| CD10                           | HI10a     |
| CD101 (BB27)                   | BB27      |
| CD103 (Integrin $\alpha$ E)    | Ber-ACT8  |
| CD106                          | STA       |
| CD107a (LAMP-1)                | H4A3      |
| CD112 (Nectin-2)               | TX31      |
| CD117 (c-kit)                  | 104D2     |
| CD11a                          | TS2/4     |
| CD11b                          | M1/70     |
| CD11c                          | S-HCL-3   |
| CD122 (IL-2R $\beta$ )         | TU27      |
| CD123                          | 6H6       |
| CD124 (IL-4R $\alpha$ )        | G077F6    |
| CD127 (IL-7R $\alpha$ )        | A019D5    |
| CD133                          | S16016B   |
| CD137 (4-1BB)                  | 4B4-1     |
| CD137L (4-1BB Ligand)          | 5F4       |
| CD138 (Syndecan-1)             | DL-101    |
| CD14                           | M5E2      |
| CD141 (Thrombomodulin)         | M80       |
| CD144 (VE-Cadherin)            | BV9       |
| CD146                          | P1H12     |
| CD15 (SSEA-1)                  | W6D3      |
| CD150 (SLAM)                   | A12 (7D4) |
| CD152 (CTLA-4)                 | BNI3      |
| CD154                          | 24-31     |
| CD155 (PVR)                    | SKII.4    |
| CD158 (KIR2DL1/S1/S3/S5)       | HP-MA4    |
| CD158b (KIR2DL2/L3, NKAT2)     | DX27      |
| CD158e1 (KIR3DL1, NKB1)        | DX9       |
| CD158f (KIR2DL5)               | UP-R1     |
| CD16                           | 3G8       |
| CD161                          | HP-3G10   |
| CD163                          | GHI/61    |
| CD169 (Sialoadhesin, Siglec-1) | 7-239     |
| CD178 (Fas-L)                  | NOK-1     |
| CD18                           | TS1/18    |
| CD183 (CXCR3)                  | G025H7    |
| CD184 (CXCR4)                  | 12G5      |
| CD185 (CXCR5)                  | J252D4    |
| CD19                           | HIB19     |
| CD193 (CCR3)                   | 5E8       |
| CD194 (CCR4)                   | L291H4    |

Table A.1 continued from previous page

| Surface protein       | Clone     |
|-----------------------|-----------|
| CD195 (CCR5)          | J418F1    |
| CD196 (CCR6)          | G034E3    |
| CD197 (CCR7)          | G043H7    |
| CD1a                  | HI149     |
| CD1c                  | L161      |
| CD1d                  | 51.1      |
| CD2                   | TS1/8     |
| CD20                  | 2H7       |
| CD204                 | 7C9C20    |
| CD206 (MMR)           | 15-2      |
| CD207                 | 4C7       |
| CD209 (DC-SIGN)       | 9E9A8     |
| CD21                  | Bu32      |
| CD22                  | S-HCL-1   |
| CD223 (LAG-3)         | 11C3C65   |
| CD224                 | KF29      |
| CD226 (DNAM-1)        | 11A8      |
| CD23                  | EBVCS-5   |
| CD235ab               | HIR2      |
| CD24                  | ML5       |
| CD244 (2B4)           | C1.7      |
| CD25                  | BC96      |
| CD252 (OX40L)         | 11C3.1    |
| CD254 (TRANCE, RANKL) | MIH24     |
| CD257 (BAFF, BLYS)    | 1D6       |
| CD258 (LIGHT)         | T5-39     |
| CD26                  | BA5b      |
| CD267 (TACI)          | 1A1       |
| CD268 (BAFF-R)        | 11C1      |
| CD269 (BCMA)          | 19F2      |
| CD27                  | O323      |
| CD272 (BTLA)          | MIH26     |
| CD273 (B7-DC, PD-L2)  | 24F.10C12 |
| CD274 (B7-H1, PD-L1)  | 29E.2A3   |
| CD275 (B7-H2, ICOSL)  | 2D3       |
| CD278 (ICOS)          | C398.4A   |
| CD279 (PD-1)          | EH12.2H7  |
| CD28                  | CD28.2    |
| CD29                  | TS2/16    |
| CD294 (CRTH2)         | BM16      |
| CD3                   | UCHT1     |
| CD30                  | BY88      |
| CD303 (BDCA-2)        | 201A      |
| CD304 (Neuropilin-1)  | 12C2      |

Table A.1 continued from previous page

| Surface protein      | Clone   |
|----------------------|---------|
| CD305 (LAIR1)        | NKTA255 |
| CD307d (FcRL4)       | 413D12  |
| CD307e (FcRL5)       | 509f6   |
| CD309 (VEGFR2)       | 7D4-6   |
| CD31                 | WM59    |
| CD314 (NKG2D)        | 1D11    |
| CD319 (CRACC)        | 162.1   |
| CD32                 | FUN-2   |
| CD324 (E-Cadherin)   | 67A4    |
| CD326 (Ep-CAM)       | 9C4     |
| CD328 (Siglec-7)     | 6-434   |
| CD33                 | P67.6   |
| CD335 (NKp46)        | 9E2     |
| CD336 (NKp44)        | P44-8   |
| CD337 (NKp30)        | P30-15  |
| CD34                 | 581     |
| CD35                 | E11     |
| CD357 (GITR)         | 108-17  |
| CD36                 | 5-271   |
| CD360 (IL-21R)       | 4B2.9   |
| CD366 (Tim-3)        | F38-2E2 |
| CD370 (CLEC9A/DNGR1) | 8F9     |
| CD38                 | HIT2    |
| CD39                 | A1      |
| CD4                  | RPA-T4  |
| CD40                 | 5C3     |
| CD41                 | HIP8    |
| CD44                 | IM7     |
| CD45                 | HI30    |
| CD45RA               | HI100   |
| CD45RO               | UCHL1   |
| CD47                 | CC2C6   |
| CD49b                | P1E6-C5 |
| CD49d                | 9F10    |
| CD49f                | GoH3    |
| CD5                  | UCHT2   |
| CD52                 | HI186   |
| CD54                 | HA58    |
| CD56 (NCAM)          | 5.1H11  |
| CD57 Recombinant     | QA17A04 |
| CD58 (LFA-3)         | TS2/9   |
| CD62L                | DREG-56 |
| CD62P (P-Selectin)   | AK4     |
| CD64                 | 10.1    |

Table A.1 continued from previous page

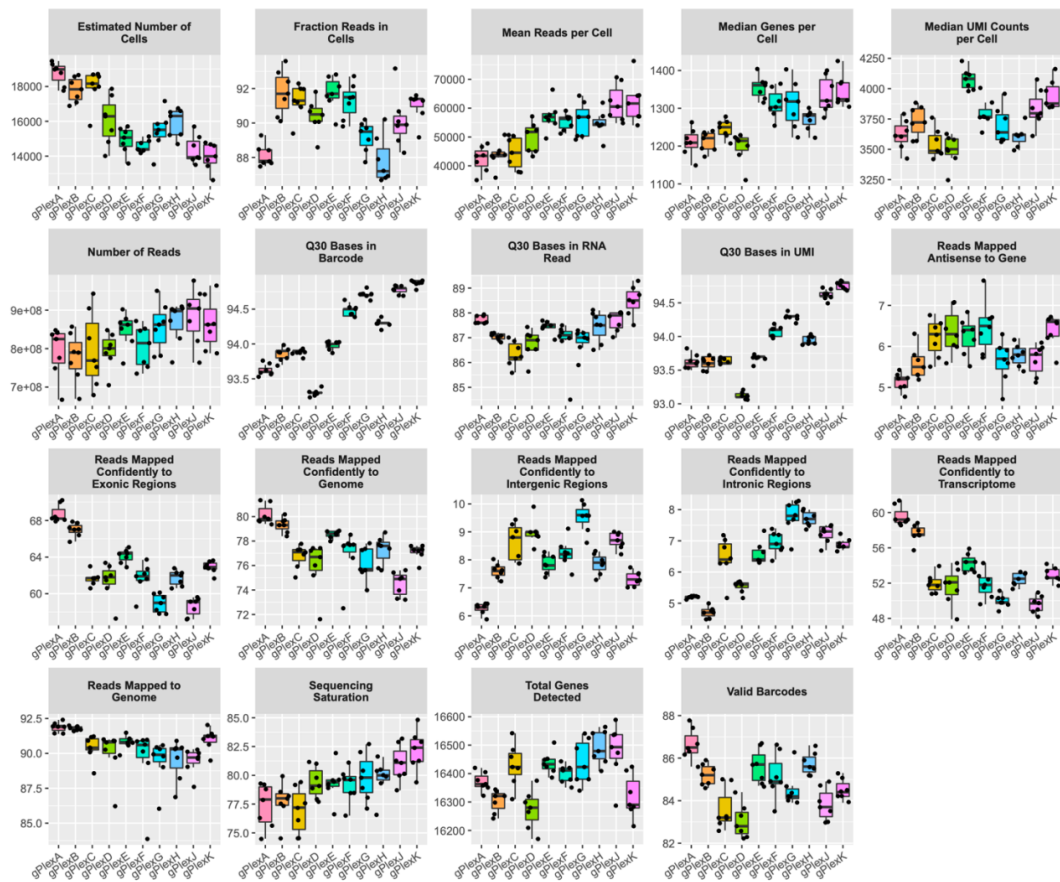
| Surface protein              | Clone          |
|------------------------------|----------------|
| CD66a/c/e                    | ASL-32         |
| CD66b                        | 6/40c          |
| CD69                         | FN50           |
| CD7                          | CD7-6B7        |
| CD70                         | 113-16         |
| CD71                         | CY1G4          |
| CD73 (Ecto-5'-nucleotidase)  | AD2            |
| CD79b (Ig $\beta$ )          | CB3-1          |
| CD8                          | SK1            |
| CD80                         | 2D10           |
| CD81 (TAPA-1)                | 5A6            |
| CD82                         | ASL-24         |
| CD83                         | HB15e          |
| CD85j (ILT2)                 | GHI/75         |
| CD86                         | IT2.2          |
| CD88 (C5aR)                  | S5/1           |
| CD90 (Thy1)                  | 5E10           |
| CD94                         | DX22           |
| CD95 (Fas)                   | DX2            |
| CD96 (TACTILE)               | NK92.39        |
| CD98                         | MEM-108        |
| CD99                         | 3B2/TA8        |
| CLEC12A                      | 50C1           |
| c-Met                        | 12.1           |
| CX3CR1                       | K0124E1        |
| DR3 (TRAMP)                  | JD3            |
| Fc $\epsilon$ RI $\alpha$    | AER-37 (CRA-1) |
| GARP (LRRC32)                | 7B11           |
| HLA-A,B,C                    | W6/32          |
| HLA-A2                       | BB7.2          |
| HLA-DR                       | L243           |
| HLA-F                        | 3D11/HLA-F     |
| Ig light chain $\kappa$      | MHK-49         |
| Ig light chain $\lambda$     | MHL-38         |
| IgA                          | HP6123         |
| IgD                          | IA6-2          |
| IgG Fc                       | M1310G05       |
| IgG1, $\kappa$ isotype Ctrl  | MOPC-21        |
| IgG2a, $\kappa$ isotype Ctrl | MOPC-173       |
| IgG2b, $\kappa$ isotype Ctrl | MPC-11         |
| IgG2b, $\kappa$ Isotype Ctrl | RTK4530        |
| IgM                          | MHM-88         |
| integrin $\beta$ 7           | FIB504         |
| KLRG1 (MAFA)                 | SA231A2        |



**Table A.1 continued from previous page**

| Surface protein                             | Clone   |
|---------------------------------------------|---------|
| LOX-1                                       | 15C4    |
| Mac-2 (Galectin-3)                          | M3/38   |
| NLRP2                                       | 8F10B51 |
| Podocalyxin                                 | mAb 84  |
| Podoplanin                                  | NC-08   |
| TCR V $\alpha$ 24-J $\alpha$ 18 (iNKT cell) | 6B11    |
| TCR V $\alpha$ 7.2                          | 3C10    |
| TCR V $\beta$ 13.1                          | H131    |
| TCR V $\gamma$ 9                            | B3      |
| TCR V $\delta$ 2                            | B6      |
| TCR $\alpha/\beta$                          | IP26    |
| TCR $\gamma/\delta$                         | B1      |
| TIGIT (VSTM3)                               | A15153G |
| TSLPR (TSLP-R)                              | 1D3     |
| XCR1                                        | S15046E |

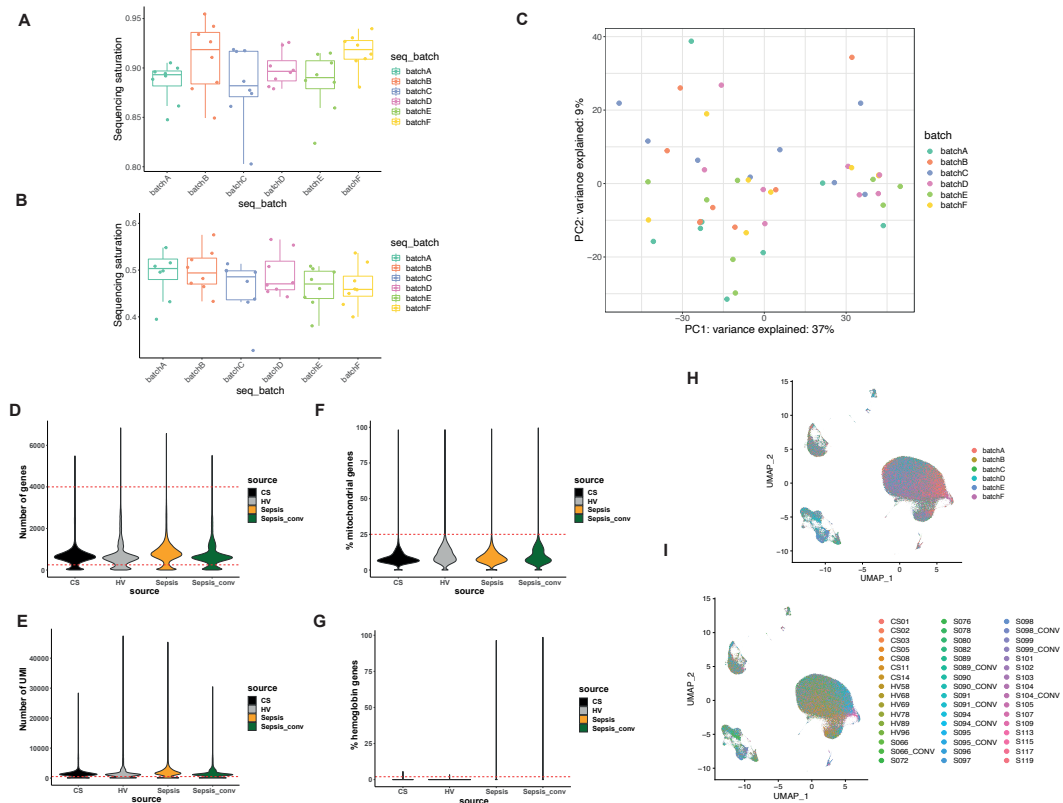
**Table A.1:** BioLegend TotalSeq-C CITE-seq antibodies.



**Figure A.2: CITE-seq sequencing quality control (QC).** Cell number capture, reads and UMI detected per cell, read mapping metrics and sequencing saturation QC metrics shown per pool (gPlex) with each channel per pool plotted.

| Surface protein | Cat. No. |
|-----------------|----------|
| CD8             | 940003   |
| CD4             | 940001   |
| PD-1 (CD279)    | 940015   |
| CD19            | 940004   |
| CD56 (NCAM-1)   | 940007   |
| HLA-DR          | 940010   |
| CD86            | 940025   |
| CD14            | 940005   |
| CD33            | 940031   |
| CD16            | 940006   |
| CD10            | 940045   |
| CD11b (Mac-1)   | 940008   |
| CD11c           | 940024   |
| CD62L           | 940041   |
| CD38            | 940013   |
| PDL1 (CD274)    | 940035   |
| CD24            | 940028   |
| CD49d           | 940059   |
| IGSF2 (CD101)   | 940269   |
| IL-3RA (CD123)  | 940020   |
| CD35            | 940248   |
| CD32            | 940069   |
| PDL2 (CD273)    | 940071   |
| CD15            | 940274   |
| CD64            | 940023   |
| CD71            | 940275   |
| CD63            | 940243   |
| CD371           | 940212   |
| CD43            | 940278   |
| CD163           | 940058   |

**Table A.2:** BD Rhapsody Abseq antibodies.



**Figure A.3: scRNA-seq + Abseq quality control (QC).** (A-B) Sequencing saturation for (A) single cell RNA and (B) protein libraries per sequencing batch (seq batch). (C) First two principal components (PCs) from principal components analysis (PCA) of all samples of cohort 1 pseudobulked colored by sequencing batch. (D-G) QC violin plots for various filtering thresholds (red dotted lines) for (D) number of genes detected per cell (E) number of UMIs per cell (F) percentage of mitochondrial genes detected per cell and (G) percentage of hemoglobin genes detected per cell. (H-I) UMAP of 272,993 cells after QC filtering colored by (H) sequencing batch or (I) sample identifier.

# B

## APPENDIX TO CHAPTER 3

| Gene              | Log2 Fold Change | P Value  | FDR      |
|-------------------|------------------|----------|----------|
| <i>IL2RB</i>      | -1.66            | 9.71E-11 | 4.60E-08 |
| <i>ETS2</i>       | 1.48             | 9.89E-11 | 4.60E-08 |
| <i>ZNF445</i>     | 1.11             | 1.01E-10 | 4.60E-08 |
| <i>ZDHHC3</i>     | 0.96             | 1.01E-10 | 4.60E-08 |
| <i>PFKFB2</i>     | 2.83             | 1.02E-10 | 4.60E-08 |
| <i>TRIM44</i>     | -1.08            | 8.86E-11 | 4.44E-08 |
| <i>KIF1B</i>      | 1.21             | 8.64E-11 | 4.43E-08 |
| <i>PGD</i>        | 1.26             | 8.36E-11 | 4.40E-08 |
| <i>JAK3</i>       | 1.26             | 8.40E-11 | 4.40E-08 |
| <i>ALOX5</i>      | 1.27             | 7.99E-11 | 4.39E-08 |
| <i>ANKS1A</i>     | 1.44             | 6.66E-11 | 3.74E-08 |
| <i>BLOC1S1</i>    | 0.92             | 6.39E-11 | 3.68E-08 |
| <i>GOLGA1</i>     | 1.29             | 5.37E-11 | 3.17E-08 |
| <i>CD63</i>       | 1.07             | 5.21E-11 | 3.16E-08 |
| <i>TRAC</i>       | -2.3             | 4.80E-11 | 3.06E-08 |
| <i>SYNJ2BP</i>    | -1.26            | 4.92E-11 | 3.06E-08 |
| <i>TRBC1</i>      | -1.9             | 4.45E-11 | 2.97E-08 |
| <i>ST6GALNAC3</i> | 1.77             | 4.51E-11 | 2.97E-08 |
| <i>LAIR1</i>      | 1.71             | 4.03E-11 | 2.82E-08 |
| <i>GPR174</i>     | -1.9             | 3.79E-11 | 2.73E-08 |
| <i>RGL4</i>       | 1.93             | 3.33E-11 | 2.53E-08 |
| <i>CD2</i>        | -1.89            | 3.37E-11 | 2.53E-08 |
| <i>CD74</i>       | -1.58            | 3.40E-11 | 2.53E-08 |
| <i>CD3E</i>       | -2.15            | 2.84E-11 | 2.34E-08 |
| <i>DDAH2</i>      | 1.33             | 2.37E-11 | 2.03E-08 |
| <i>NCR3</i>       | -1.22            | 2.23E-11 | 2.02E-08 |
| <i>BCL9L</i>      | -1.38            | 2.28E-11 | 2.02E-08 |
| <i>KLF7</i>       | 1.13             | 1.52E-11 | 1.72E-08 |
| <i>RASGRP1</i>    | -1.52            | 1.73E-11 | 1.72E-08 |
| <i>LINC02273</i>  | -1.11            | 1.74E-11 | 1.72E-08 |
| <i>DLG3</i>       | -1.15            | 1.74E-11 | 1.72E-08 |
| <i>STXBP2</i>     | 1.17             | 1.79E-11 | 1.72E-08 |
| <i>GIMAP7</i>     | -1.78            | 1.79E-11 | 1.72E-08 |
| <i>ACVR1B</i>     | 1.26             | 1.25E-11 | 1.60E-08 |
| <i>FCAR</i>       | 1.61             | 1.01E-11 | 1.50E-08 |
| <i>GADD45A</i>    | 2.2              | 1.03E-11 | 1.50E-08 |
| <i>HLA-DMB</i>    | -1.93            | 1.07E-11 | 1.50E-08 |
| <i>HLA-DPB1</i>   | -1.95            | 1.10E-11 | 1.50E-08 |

Table B.1 continued from previous page

| Gene           | Log2 Fold Change | P Value  | FDR      |
|----------------|------------------|----------|----------|
| <i>TIGIT</i>   | -1.39            | 6.45E-12 | 1.14E-08 |
| <i>GATA3</i>   | -1.47            | 5.03E-12 | 9.66E-09 |
| <i>TRDC</i>    | -1.63            | 1.79E-12 | 4.49E-09 |
| <i>CD8B</i>    | -1.74            | 2.04E-12 | 4.49E-09 |
| <i>ARL4C</i>   | -1.82            | 2.13E-12 | 4.49E-09 |
| <i>SYNE1</i>   | 1.25             | 2.14E-12 | 4.49E-09 |
| <i>UPP1</i>    | 1.68             | 6.47E-13 | 2.18E-09 |
| <i>AMOT</i>    | -1.12            | 6.62E-13 | 2.18E-09 |
| <i>CD82</i>    | 1.48             | 1.40E-13 | 1.06E-09 |
| <i>NSUN7</i>   | 1.75             | 2.24E-13 | 1.06E-09 |
| <i>OPLAH</i>   | 2.29             | 2.30E-13 | 1.06E-09 |
| <i>GZMK</i>    | -1.74            | 2.43E-14 | 5.19E-10 |
| <i>RABGEF1</i> | 1.1              | 4.50E-14 | 5.19E-10 |

**Table B.1: Summary of genes differentially expressed between SRS (all-cause sepsis).**  
50 most significantly differentially expressed genes.

| Gene           | Log2 Fold Change | P Value  | FDR      |
|----------------|------------------|----------|----------|
| <i>SPI1</i>    | 2.21             | 4.02E-23 | 1.85E-20 |
| <i>ARID3A</i>  | 1.63             | 3.77E-23 | 1.79E-20 |
| <i>S100A8</i>  | 3.15             | 3.80E-23 | 1.79E-20 |
| <i>MGAM2</i>   | 4.18             | 3.60E-23 | 1.77E-20 |
| <i>ALOX5</i>   | 2.6              | 2.32E-23 | 1.16E-20 |
| <i>TDRD9</i>   | 4.13             | 2.20E-23 | 1.13E-20 |
| <i>NBEAL2</i>  | 2.21             | 2.21E-23 | 1.13E-20 |
| <i>AGTRAP</i>  | 1.97             | 1.70E-23 | 9.14E-21 |
| <i>RPL34</i>   | -1.89            | 1.59E-23 | 8.74E-21 |
| <i>PADI4</i>   | 2.57             | 1.49E-23 | 8.57E-21 |
| <i>DOK3</i>    | 2.31             | 1.52E-23 | 8.57E-21 |
| <i>S100A9</i>  | 2.84             | 8.25E-24 | 4.88E-21 |
| <i>NLRC4</i>   | 2.63             | 8.02E-24 | 4.87E-21 |
| <i>JAK3</i>    | 2.34             | 7.39E-24 | 4.61E-21 |
| <i>PFKFB3</i>  | 3.78             | 6.02E-24 | 3.86E-21 |
| <i>FCAR</i>    | 3.26             | 5.58E-24 | 3.68E-21 |
| <i>MXD3</i>    | 2.46             | 5.29E-24 | 3.67E-21 |
| <i>CCDC65</i>  | -1.86            | 5.42E-24 | 3.67E-21 |
| <i>KLRB1</i>   | -3.17            | 4.54E-24 | 3.28E-21 |
| <i>USB1</i>    | 2.2              | 4.13E-24 | 3.07E-21 |
| <i>ACAA1</i>   | 1.61             | 3.37E-24 | 2.59E-21 |
| <i>MMP9</i>    | 4.79             | 1.94E-24 | 1.54E-21 |
| <i>SEMA4A</i>  | 2.27             | 1.47E-24 | 1.21E-21 |
| <i>CD177</i>   | 8.56             | 1.12E-24 | 9.57E-22 |
| <i>PRKCD</i>   | 1.76             | 9.63E-25 | 8.55E-22 |
| <i>SLC26A8</i> | 3.45             | 8.69E-25 | 8.02E-22 |
| <i>KIF1B</i>   | 2.68             | 5.35E-25 | 5.14E-22 |
| <i>BCL6</i>    | 3.14             | 3.46E-25 | 3.47E-22 |

Table B.2 continued from previous page

| Gene              | Log2 Fold Change | P Value  | FDR      |
|-------------------|------------------|----------|----------|
| <i>OPLAH</i>      | 4.4              | 2.33E-25 | 2.54E-22 |
| <i>AC023590.1</i> | -0.63            | 2.43E-25 | 2.54E-22 |
| <i>MAPK14</i>     | 2.83             | 1.70E-25 | 1.96E-22 |
| <i>ITGAM</i>      | 2.39             | 1.34E-25 | 1.63E-22 |
| <i>CEP126</i>     | -1.61            | 9.53E-26 | 1.22E-22 |
| <i>FCER1A</i>     | -3.63            | 4.36E-26 | 5.92E-23 |
| <i>SCARA5</i>     | -0.75            | 3.55E-26 | 5.12E-23 |
| <i>PRSS22</i>     | -0.6             | 3.24E-26 | 4.98E-23 |
| <i>LTB4R</i>      | 2.26             | 2.50E-26 | 4.48E-23 |
| <i>MAFG</i>       | 2.04             | 2.68E-26 | 4.48E-23 |
| <i>DNASE1L3</i>   | -1.02            | 2.72E-26 | 4.48E-23 |
| <i>SEZ6L</i>      | -1.72            | 1.77E-26 | 3.72E-23 |
| <i>MCEMP1</i>     | 4.74             | 7.69E-27 | 1.77E-23 |
| <i>ANXA3</i>      | 4.12             | 5.17E-27 | 1.33E-23 |
| <i>DYSF</i>       | 3.39             | 8.07E-28 | 2.33E-24 |
| <i>GALNT14</i>    | 4.36             | 5.96E-28 | 1.96E-24 |
| <i>AL357143.1</i> | -1.72            | 4.82E-28 | 1.85E-24 |
| <i>OSCAR</i>      | 2.6              | 1.37E-28 | 6.34E-25 |
| <i>HK3</i>        | 3.46             | 9.11E-29 | 5.25E-25 |
| <i>TIFAB</i>      | -1.57            | 1.51E-30 | 1.16E-26 |
| <i>COL26A1</i>    | -0.98            | 2.98E-31 | 3.90E-27 |
| <i>ENHO</i>       | -1.13            | 3.38E-31 | 3.90E-27 |
| <i>RABGEF1</i>    | 1.1              | 4.50E-14 | 5.19E-10 |

Table B.2: Summary of genes differentially expressed between sepsis-HV. 50 most significantly differentially expressed genes.

| Gene            | Log2 Fold Change | P Value  | FDR      |
|-----------------|------------------|----------|----------|
| <i>GPR132</i>   | 0.71             | 9.29E-12 | 3.76E-09 |
| <i>STIMATE</i>  | 1.03             | 9.09E-12 | 3.75E-09 |
| <i>CDC14B</i>   | -1.34            | 7.54E-12 | 3.16E-09 |
| <i>KIAA0930</i> | 0.88             | 7.22E-12 | 3.08E-09 |
| <i>STMN1</i>    | -1.69            | 6.66E-12 | 2.90E-09 |
| <i>CRACD</i>    | -1.38            | 6.12E-12 | 2.77E-09 |
| <i>JCHAIN</i>   | -3.43            | 6.24E-12 | 2.77E-09 |
| <i>PHKA2</i>    | 0.79             | 5.46E-12 | 2.52E-09 |
| <i>LAIR1</i>    | 1.35             | 4.82E-12 | 2.27E-09 |
| <i>FUT8</i>     | -1.08            | 4.69E-12 | 2.25E-09 |
| <i>ZNF664</i>   | -1.03            | 3.44E-12 | 1.69E-09 |
| <i>SPSB1</i>    | -1.1             | 3.13E-12 | 1.57E-09 |
| <i>NCKAP5L</i>  | 0.89             | 2.81E-12 | 1.44E-09 |
| <i>FAM160A2</i> | 0.98             | 2.41E-12 | 1.26E-09 |
| <i>LENG9</i>    | 0.86             | 2.28E-12 | 1.23E-09 |
| <i>LIPH</i>     | -1.45            | 2.08E-12 | 1.14E-09 |
| <i>TAZ</i>      | 0.76             | 1.87E-12 | 1.09E-09 |
| <i>CDK6</i>     | -0.95            | 1.91E-12 | 1.09E-09 |

Table B.3 continued from previous page

| Gene              | Log2 Fold Change | P Value  | FDR      |
|-------------------|------------------|----------|----------|
| <i>LINC00513</i>  | 1.41             | 1.94E-12 | 1.09E-09 |
| <i>AC004771.4</i> | 0.6              | 1.18E-12 | 9.53E-10 |
| <i>AC016831.5</i> | 1.29             | 1.18E-12 | 9.53E-10 |
| <i>RUNX2</i>      | -0.88            | 1.25E-12 | 9.53E-10 |
| <i>DNASE1L1</i>   | 0.74             | 1.33E-12 | 9.53E-10 |
| <i>SEPTIN11</i>   | -1.09            | 1.38E-12 | 9.53E-10 |
| <i>IGHG1</i>      | -4.35            | 1.40E-12 | 9.53E-10 |
| <i>ARRDC1</i>     | 0.73             | 1.41E-12 | 9.53E-10 |
| <i>ABI2</i>       | -1.03            | 1.43E-12 | 9.53E-10 |
| <i>SLA2</i>       | -1.2             | 1.45E-12 | 9.53E-10 |
| <i>DENND4B</i>    | 0.68             | 9.22E-13 | 8.18E-10 |
| <i>ITGA1</i>      | -1.13            | 6.57E-13 | 6.06E-10 |
| <i>IGLV3-25</i>   | -3.45            | 5.64E-13 | 5.42E-10 |
| <i>SCD</i>        | -1.79            | 5.11E-13 | 5.12E-10 |
| <i>TUBGCP4</i>    | -0.88            | 3.73E-13 | 3.91E-10 |
| <i>MCM6</i>       | -1.37            | 2.24E-13 | 2.72E-10 |
| <i>SNX15</i>      | 0.66             | 1.98E-13 | 2.54E-10 |
| <i>AF064858.1</i> | 2.69             | 1.35E-13 | 1.82E-10 |
| <i>MNT</i>        | 0.8              | 5.34E-14 | 8.22E-11 |
| <i>MAPK8IP3</i>   | 1.11             | 3.89E-14 | 6.41E-11 |
| <i>ACAA1</i>      | 0.86             | 2.89E-14 | 5.49E-11 |
| <i>LINC00884</i>  | 1.69             | 3.02E-14 | 5.49E-11 |
| <i>NIPA1</i>      | -1.2             | 3.10E-14 | 5.49E-11 |
| <i>TNIK</i>       | -1               | 2.07E-14 | 5.20E-11 |
| <i>LRBA</i>       | -0.91            | 2.26E-14 | 5.20E-11 |
| <i>GBGT1</i>      | 1.31             | 1.44E-14 | 4.15E-11 |
| <i>PALM2AKAP2</i> | -1.88            | 9.54E-15 | 3.14E-11 |
| <i>PPM1M</i>      | 1.15             | 4.54E-15 | 1.75E-11 |
| <i>KLF14</i>      | 1.81             | 3.63E-15 | 1.67E-11 |
| <i>KRT8P12</i>    | 1.16             | 5.91E-16 | 3.41E-12 |
| <i>COLGALT1</i>   | 1.16             | 2.42E-16 | 2.38E-12 |
| <i>NXF1</i>       | 0.62             | 3.10E-16 | 2.38E-12 |
| <i>RABGEF1</i>    | 1.1              | 4.50E-14 | 5.19E-10 |

Table B.3: Summary of genes differentially expressed between all-cause vs. COVID-19 sepsis. 50 most significantly differentially expressed genes.

| Gene              | Log2 Fold Change | P value  | FDR      |
|-------------------|------------------|----------|----------|
| <i>CHIT1</i>      | 2.03             | 2.00E-09 | 8.87E-07 |
| <i>AL133517.1</i> | 1                | 1.87E-09 | 8.46E-07 |
| <i>ABHD4</i>      | 0.84             | 1.69E-09 | 7.78E-07 |
| <i>CLU</i>        | 1.4              | 1.49E-09 | 7.06E-07 |
| <i>ELAC1</i>      | -0.72            | 1.50E-09 | 7.06E-07 |
| <i>BMX</i>        | 1.84             | 1.21E-09 | 6.03E-07 |
| <i>PGD</i>        | 1.22             | 1.23E-09 | 6.03E-07 |
| <i>PADI4</i>      | 1.29             | 1.09E-09 | 5.57E-07 |



Table B.4 continued from previous page

| Gene              | Log2 Fold Change | P value  | FDR      |
|-------------------|------------------|----------|----------|
| <i>MAPK14</i>     | 1.34             | 9.99E-10 | 5.24E-07 |
| <i>SLC37A3</i>    | 1.34             | 9.68E-10 | 5.19E-07 |
| <i>PTPRJ</i>      | 0.92             | 9.04E-10 | 4.97E-07 |
| <i>SRGN</i>       | 1.08             | 7.47E-10 | 4.33E-07 |
| <i>GBE1</i>       | 0.79             | 7.51E-10 | 4.33E-07 |
| <i>RNF24</i>      | 1.19             | 6.89E-10 | 4.18E-07 |
| <i>TMEM116</i>    | -0.68            | 6.55E-10 | 4.08E-07 |
| <i>ANXA3</i>      | 1.9              | 5.72E-10 | 3.85E-07 |
| <i>PRKCB</i>      | 0.63             | 5.90E-10 | 3.85E-07 |
| <i>HP</i>         | 2.42             | 6.02E-10 | 3.85E-07 |
| <i>ARHGEF9</i>    | -0.87            | 5.49E-10 | 3.84E-07 |
| <i>ABCA13</i>     | 2.64             | 5.28E-10 | 3.80E-07 |
| <i>HDAC4</i>      | 1.08             | 4.80E-10 | 3.57E-07 |
| <i>CCDC65</i>     | -0.94            | 4.55E-10 | 3.50E-07 |
| <i>CRTAM</i>      | -1.2             | 4.29E-10 | 3.41E-07 |
| <i>AL359220.1</i> | -0.89            | 3.83E-10 | 3.16E-07 |
| <i>TRGV9</i>      | -1.11            | 3.57E-10 | 3.08E-07 |
| <i>PLD1</i>       | 0.89             | 3.61E-10 | 3.08E-07 |
| <i>LINC02084</i>  | -0.94            | 3.12E-10 | 2.88E-07 |
| <i>DYSF</i>       | 1.56             | 2.57E-10 | 2.55E-07 |
| <i>CTSA</i>       | 0.95             | 2.58E-10 | 2.55E-07 |
| <i>CUX1</i>       | 0.84             | 2.65E-10 | 2.55E-07 |
| <i>DACH1</i>      | 1.58             | 1.30E-10 | 1.49E-07 |
| <i>CLTCL1</i>     | 1.34             | 1.36E-10 | 1.49E-07 |
| <i>MSRB3</i>      | 1.22             | 8.94E-11 | 1.44E-07 |
| <i>MCU</i>        | 0.78             | 1.05E-10 | 1.44E-07 |
| <i>AL139424.2</i> | 1.29             | 1.12E-10 | 1.44E-07 |
| <i>CPNE2</i>      | 1                | 1.18E-10 | 1.44E-07 |
| <i>ITGAM</i>      | 1.2              | 1.19E-10 | 1.44E-07 |
| <i>GSN</i>        | 1.23             | 7.49E-11 | 1.33E-07 |
| <i>TLN1</i>       | 1.03             | 6.25E-11 | 1.30E-07 |
| <i>KIF1B</i>      | 1.39             | 6.75E-11 | 1.30E-07 |
| <i>IGF1R</i>      | 1.31             | 5.59E-11 | 1.29E-07 |
| <i>ACTN1</i>      | 1.07             | 3.11E-11 | 7.96E-08 |
| <i>PYGL</i>       | 1.56             | 1.56E-11 | 5.90E-08 |
| <i>BCORL1</i>     | 0.97             | 1.73E-11 | 5.90E-08 |
| <i>APMAP</i>      | 1.29             | 1.77E-11 | 5.90E-08 |
| <i>ANO10</i>      | 0.99             | 1.96E-11 | 5.90E-08 |
| <i>MED12L</i>     | 1.65             | 2.05E-11 | 5.90E-08 |
| <i>BCL6</i>       | 1.73             | 5.09E-12 | 3.91E-08 |
| <i>PGLYRP1</i>    | 2.15             | 6.07E-13 | 1.14E-08 |
| <i>MAFG</i>       | 1.13             | 9.92E-13 | 1.14E-08 |

Table B.4: Summary of genes differentially expressed between critical vs. mild COVID-19. 50 most significantly differentially expressed genes.

## APPENDIX TO CHAPTER 4

## Composition analysis of minor cell subsets

Frequency of minor subsets out of total PBMCs by clinical category

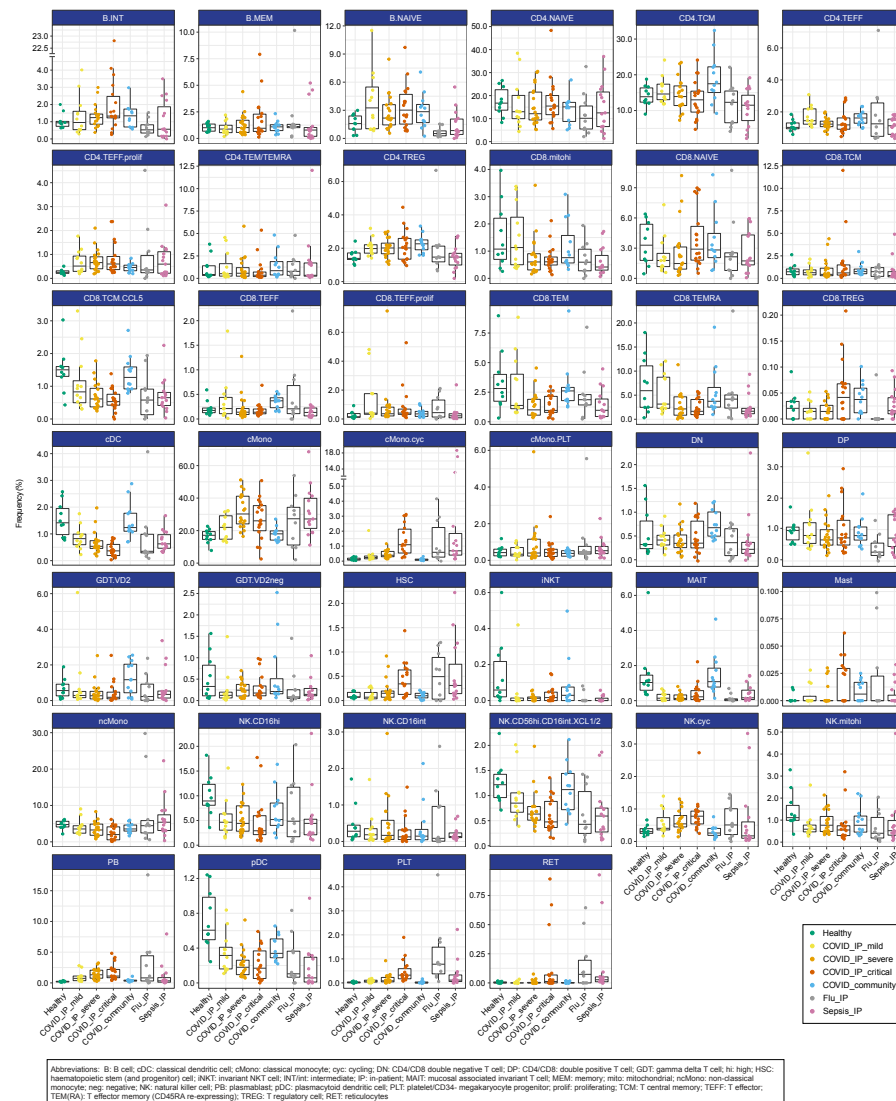


Figure C.1: CITE-seq minor cell subset proportions across comparator groups.

| Module       | Gene           | Correlation with eigengene |
|--------------|----------------|----------------------------|
| brown        | <i>DDX5</i>    | 0.749                      |
| brown        | <i>CNOT7</i>   | 0.747                      |
| brown        | <i>HNRNPM</i>  | 0.727                      |
| brown        | <i>GPBP1</i>   | 0.727                      |
| brown        | <i>UBQLN2</i>  | 0.724                      |
| greenyellow  | <i>PPIB</i>    | 0.948                      |
| greenyellow  | <i>HSP90B1</i> | 0.938                      |
| greenyellow  | <i>MYDGF</i>   | 0.936                      |
| greenyellow  | <i>P4HB</i>    | 0.918                      |
| greenyellow  | <i>MZB1</i>    | 0.916                      |
| black        | <i>COX6A1</i>  | 0.837                      |
| black        | <i>COX7B</i>   | 0.811                      |
| black        | <i>MIF</i>     | 0.804                      |
| black        | <i>PSMB3</i>   | 0.8                        |
| black        | <i>ATP5F1C</i> | 0.785                      |
| cyan         | <i>RRM2</i>    | 0.841                      |
| cyan         | <i>ZWINT</i>   | 0.833                      |
| cyan         | <i>CCNA2</i>   | 0.799                      |
| cyan         | <i>CDC45</i>   | 0.795                      |
| cyan         | <i>UBE2C</i>   | 0.787                      |
| magenta      | <i>RPS15A</i>  | 0.941                      |
| magenta      | <i>EEF1A1</i>  | 0.929                      |
| magenta      | <i>RPS21</i>   | 0.92                       |
| magenta      | <i>RPL18</i>   | 0.906                      |
| magenta      | <i>RPL19</i>   | 0.904                      |
| midnightblue | <i>NOP10</i>   | 0.828                      |
| midnightblue | <i>HSPD1</i>   | 0.815                      |
| midnightblue | <i>CCT2</i>    | 0.811                      |
| midnightblue | <i>HAX1</i>    | 0.81                       |
| midnightblue | <i>COPB2</i>   | 0.803                      |
| turquoise    | <i>MAT2B</i>   | 0.814                      |
| turquoise    | <i>TNFAIP8</i> | 0.787                      |
| turquoise    | <i>BLNK</i>    | 0.771                      |
| turquoise    | <i>DCK</i>     | 0.763                      |
| turquoise    | <i>WASHC4</i>  | 0.753                      |
| blue         | <i>DAZAP2</i>  | 0.91                       |
| blue         | <i>SF1</i>     | 0.907                      |
| blue         | <i>KMT2E</i>   | 0.885                      |
| blue         | <i>EPC1</i>    | 0.881                      |
| blue         | <i>SRRM1</i>   | 0.873                      |
| pink         | <i>CLIC1</i>   | 0.779                      |
| pink         | <i>SCIMP</i>   | 0.775                      |
| pink         | <i>GRN</i>     | 0.775                      |
| pink         | <i>CALM2</i>   | 0.772                      |
| pink         | <i>ARPC5</i>   | 0.77                       |
| green        | <i>CDCA7L</i>  | 0.8                        |

Table C.1 continued from previous page

| Module | Gene           | Correlation with eigengene |
|--------|----------------|----------------------------|
| green  | <i>IL4R</i>    | 0.788                      |
| green  | <i>BACH2</i>   | 0.788                      |
| green  | <i>SNX29</i>   | 0.774                      |
| green  | <i>ITPR1</i>   | 0.772                      |
| purple | <i>MX1</i>     | 0.899                      |
| purple | <i>XAF1</i>    | 0.891                      |
| purple | <i>EPSTI1</i>  | 0.883                      |
| purple | <i>IFITM1</i>  | 0.88                       |
| purple | <i>IRF7</i>    | 0.878                      |
| grey   | <i>COL9A3</i>  | 0.547                      |
| grey   | <i>EPN1</i>    | 0.484                      |
| grey   | <i>USP22</i>   | 0.462                      |
| grey   | <i>ARMC5</i>   | 0.448                      |
| grey   | <i>FAM104B</i> | 0.44                       |

**Table C.1: Summary of B cell WGCNA module genes.** Top 5 genes per module by correlation with module eigengene expression.

| Module       | Gene              | Correlation with eigengene |
|--------------|-------------------|----------------------------|
| grey         | <i>AGAP4</i>      | 0.536                      |
| grey         | <i>CACNA1I</i>    |                            |
| grey         | <i>CCDC180</i>    | 0.493                      |
| grey         | <i>ZNF517</i>     | 0.483                      |
| grey         | <i>AL590560.2</i> | 0.476                      |
| turquoise    | <i>SPOP</i>       | 0.941                      |
| turquoise    | <i>DPF2</i>       | 0.929                      |
| turquoise    | <i>MAT2B</i>      | 0.923                      |
| turquoise    | <i>VTA1</i>       | 0.923                      |
| turquoise    | <i>KRCC1</i>      | 0.917                      |
| midnightblue | <i>RPL19</i>      | 0.981                      |
| midnightblue | <i>RPL18</i>      | 0.975                      |
| midnightblue | <i>RPL5</i>       | 0.965                      |
| midnightblue | <i>RPL22</i>      | 0.965                      |
| midnightblue | <i>RPS15A</i>     | 0.964                      |
| black        | <i>PSMD8</i>      | 0.961                      |
| black        | <i>PPM1G</i>      | 0.952                      |
| black        | <i>SKA2</i>       | 0.95                       |
| black        | <i>PSMB2</i>      | 0.945                      |
| black        | <i>ATP5MF</i>     | 0.944                      |
| green        | <i>KMT2E</i>      | 0.935                      |
| green        | <i>DAZAP2</i>     | 0.918                      |
| green        | <i>TTC17</i>      | 0.917                      |
| green        | <i>EPC1</i>       | 0.907                      |
| green        | <i>SRRM1</i>      | 0.904                      |
| cyan         | <i>MYO1F</i>      | 0.888                      |
| cyan         | <i>TLN1</i>       | 0.873                      |

Table C.2 continued from previous page

| Module      | Gene            | Correlation with eigengene |
|-------------|-----------------|----------------------------|
| cyan        | <i>EPS15</i>    | 0.859                      |
| cyan        | <i>EFHD2</i>    | 0.839                      |
| cyan        | <i>RABGAP1L</i> | 0.838                      |
| pink        | <i>RAB8A</i>    | 0.9                        |
| pink        | <i>CD164</i>    | 0.899                      |
| pink        | <i>TRANK1</i>   | 0.898                      |
| pink        | <i>SAMD9</i>    | 0.893                      |
| pink        | <i>SNX6</i>     | 0.887                      |
| lightgreen  | <i>JCHAIN</i>   | 0.934                      |
| lightgreen  | <i>IGKC</i>     | 0.912                      |
| lightgreen  | <i>IGKV3-20</i> | 0.904                      |
| lightgreen  | <i>IGKV1-5</i>  | 0.865                      |
| lightgreen  | <i>IGLV2-14</i> | 0.861                      |
| lightyellow | <i>SBNO2</i>    | 0.859                      |
| lightyellow | <i>XRN2</i>     | 0.856                      |
| lightyellow | <i>JAK3</i>     | 0.855                      |
| lightyellow | <i>REXO2</i>    | 0.848                      |
| lightyellow | <i>PIM1</i>     | 0.833                      |

**Table C.2: Summary of CD4+ T cell WGCNA module genes.** Top 5 genes per module by correlation with module eigengene expression.

| Module    | Gene            | Correlation with eigengene |
|-----------|-----------------|----------------------------|
| magenta   | <i>TCF7</i>     | 0.938                      |
| magenta   | <i>LEF1</i>     | 0.908                      |
| magenta   | <i>SPINT2</i>   | 0.893                      |
| magenta   | <i>NT5E</i>     | 0.892                      |
| magenta   | <i>RPL32</i>    | 0.891                      |
| pink      | <i>EPS15</i>    | 0.899                      |
| pink      | <i>PTPN4</i>    | 0.891                      |
| pink      | <i>SYNE1</i>    | 0.889                      |
| pink      | <i>APMAP</i>    | 0.886                      |
| pink      | <i>AKAP13</i>   | 0.879                      |
| purple    | <i>PSMC6</i>    | 0.932                      |
| purple    | <i>HNRNPF</i>   | 0.929                      |
| purple    | <i>MRPS18C</i>  | 0.927                      |
| purple    | <i>PSMA5</i>    | 0.923                      |
| purple    | <i>NOP10</i>    | 0.921                      |
| lightcyan | <i>JCHAIN</i>   | 0.909                      |
| lightcyan | <i>IGKV3-20</i> | 0.907                      |
| lightcyan | <i>IGKC</i>     | 0.859                      |
| lightcyan | <i>IGHA1</i>    | 0.851                      |
| lightcyan | <i>IGKV1-5</i>  | 0.834                      |
| cyan      | <i>SF3B2</i>    | 0.909                      |
| cyan      | <i>CAPN2</i>    | 0.907                      |
| cyan      | <i>ACTN4</i>    | 0.901                      |

Table C.3 continued from previous page

| Module       | Gene           | Correlation with eigengene |
|--------------|----------------|----------------------------|
| cyan         | <i>CLIC1</i>   | 0.9                        |
| cyan         | <i>RAB8A</i>   | 0.892                      |
| blue         | <i>COX5A</i>   | 0.977                      |
| blue         | <i>ATP5F1C</i> | 0.964                      |
| blue         | <i>MCM7</i>    | 0.964                      |
| blue         | <i>ATP5F1B</i> | 0.962                      |
| blue         | <i>TUBA1B</i>  | 0.961                      |
| midnightblue | <i>OAS2</i>    | 0.945                      |
| midnightblue | <i>XAF1</i>    | 0.943                      |
| midnightblue | <i>SAMD9L</i>  | 0.941                      |
| midnightblue | <i>SP100</i>   | 0.939                      |
| midnightblue | <i>EIF2AK2</i> | 0.929                      |
| grey         | <i>PARP16</i>  | 0.532                      |
| grey         | <i>RUFY3</i>   | 0.479                      |
| grey         | <i>FBXO36</i>  | 0.472                      |
| grey         | <i>TRBV7-4</i> | 0.47                       |
| grey         | <i>USP40</i>   | 0.463                      |
| yellow       | <i>MAT2B</i>   | 0.885                      |
| yellow       | <i>GIMAP6</i>  | 0.881                      |
| yellow       | <i>GIMAP2</i>  | 0.877                      |
| yellow       | <i>SH2D3C</i>  | 0.863                      |
| yellow       | <i>SQSTM1</i>  | 0.856                      |
| black        | <i>DAZAP2</i>  | 0.925                      |
| black        | <i>ZNF394</i>  | 0.904                      |
| black        | <i>EPC1</i>    | 0.898                      |
| black        | <i>DDX6</i>    | 0.894                      |
| black        | <i>YPEL5</i>   | 0.874                      |

Table C.3: Summary of CD8+ T cell WGCNA module genes. Top 5 genes per module by correlation with module eigengene expression.

| Module    | Gene           | Correlation with eigengene |
|-----------|----------------|----------------------------|
| lightcyan | <i>MIDN</i>    | 0.923                      |
| lightcyan | <i>SERTAD2</i> | 0.915                      |
| lightcyan | <i>BHLHE40</i> | 0.906                      |
| lightcyan | <i>PLK3</i>    | 0.885                      |
| lightcyan | <i>BCLAF1</i>  | 0.874                      |
| grey60    | <i>MS4A4A</i>  | 0.923                      |
| grey60    | <i>AP1S1</i>   | 0.913                      |
| grey60    | <i>NKG7</i>    | 0.907                      |
| grey60    | <i>IDH1</i>    | 0.902                      |
| grey60    | <i>RNASE2</i>  | 0.892                      |
| turquoise | <i>GIMAP8</i>  | 0.925                      |
| turquoise | <i>KDM5A</i>   | 0.923                      |
| turquoise | <i>RSBN1L</i>  | 0.911                      |
| turquoise | <i>ZC3H13</i>  | 0.905                      |

Table C.4 continued from previous page

| Module     | Gene           | Correlation with eigengene |
|------------|----------------|----------------------------|
| turquoise  | <i>TRIM27</i>  | 0.899                      |
| green      | <i>NLRP1</i>   | 0.909                      |
| green      | <i>MAP3K14</i> | 0.901                      |
| green      | <i>OGFRL1</i>  | 0.89                       |
| green      | <i>RPS6KA5</i> | 0.884                      |
| green      | <i>FOXP1</i>   | 0.879                      |
| grey       | <i>PIMREG</i>  | 0.78                       |
| grey       | <i>ANLN</i>    | 0.76                       |
| grey       | <i>MCM6</i>    | 0.754                      |
| grey       | <i>NCAPH</i>   | 0.736                      |
| grey       | <i>KIF15</i>   | 0.723                      |
| tan        | <i>NELFB</i>   | 0.874                      |
| tan        | <i>RABL6</i>   | 0.868                      |
| tan        | <i>RNPEPL1</i> | 0.858                      |
| tan        | <i>GNAI2</i>   | 0.856                      |
| tan        | <i>ARPC4</i>   | 0.849                      |
| lightgreen | <i>FKBP5</i>   | 0.923                      |
| lightgreen | <i>CD163</i>   | 0.898                      |
| lightgreen | <i>PHC2</i>    | 0.893                      |
| lightgreen | <i>TPST1</i>   | 0.89                       |
| lightgreen | <i>SAP30</i>   | 0.887                      |
| black      | <i>XAF1</i>    | 0.95                       |
| black      | <i>TRIM22</i>  | 0.949                      |
| black      | <i>SAMD9L</i>  | 0.947                      |
| black      | <i>IFI35</i>   | 0.945                      |
| black      | <i>IFITM1</i>  | 0.945                      |

Table C.4: Summary of cMono WGCNA module genes. Top 5 genes per module by correlation with module eigengene expression.

| Module | Gene            | Correlation with eigengene |
|--------|-----------------|----------------------------|
| purple | <i>NRBF2</i>    | 0.777                      |
| purple | <i>OBI1</i>     | 0.776                      |
| purple | <i>SPOP</i>     | 0.774                      |
| purple | <i>TASL</i>     | 0.773                      |
| purple | <i>FBXO8</i>    | 0.758                      |
| pink   | <i>SERPINA1</i> | 0.845                      |
| pink   | <i>ICAM2</i>    | 0.789                      |
| pink   | <i>LCP2</i>     | 0.758                      |
| pink   | <i>CASP1</i>    | 0.736                      |
| pink   | <i>MYL12B</i>   | 0.722                      |
| yellow | <i>ASGR2</i>    | 0.802                      |
| yellow | <i>NISCH</i>    | 0.794                      |
| yellow | <i>IL17RA</i>   | 0.791                      |
| yellow | <i>ANG</i>      | 0.782                      |
| yellow | <i>TFEC</i>     | 0.771                      |

Table C.5 continued from previous page

| Module      | Gene              | Correlation with eigengene |
|-------------|-------------------|----------------------------|
| salmon      | <i>COX8A</i>      | 0.883                      |
| salmon      | <i>SERF2</i>      | 0.86                       |
| salmon      | <i>NDUFA1</i>     | 0.827                      |
| salmon      | <i>ATP5MPL</i>    | 0.804                      |
| salmon      | <i>UQCRQ</i>      | 0.798                      |
| black       | <i>GNAI2</i>      | 0.801                      |
| black       | <i>RAB32</i>      | 0.742                      |
| black       | <i>MAP3K3</i>     | 0.736                      |
| black       | <i>GNB2</i>       | 0.704                      |
| black       | <i>SFPQ</i>       | 0.693                      |
| blue        | <i>BIN2</i>       | 0.834                      |
| blue        | <i>SYT17</i>      | 0.818                      |
| blue        | <i>TNFRSF8</i>    | 0.806                      |
| blue        | <i>SYTL1</i>      | 0.804                      |
| blue        | <i>ADGRE3</i>     | 0.796                      |
| green       | <i>XAF1</i>       | 0.911                      |
| green       | <i>IFI35</i>      | 0.906                      |
| green       | <i>EPSTI1</i>     | 0.898                      |
| green       | <i>IFI6</i>       | 0.896                      |
| green       | <i>OASL</i>       | 0.89                       |
| grey        | <i>IGHV1-69-2</i> | 0.887                      |
| grey        | <i>IGKV6-21</i>   | 0.865                      |
| grey        | <i>IGKV1D-16</i>  | 0.857                      |
| grey        | <i>IGLV1-36</i>   | 0.791                      |
| grey        | <i>IGLV2-14</i>   | 0.604                      |
| greenyellow | <i>PTP4A2</i>     | 0.794                      |
| greenyellow | <i>SF1</i>        | 0.778                      |
| greenyellow | <i>PRR7</i>       | 0.769                      |
| greenyellow | <i>SERTAD3</i>    | 0.756                      |
| greenyellow | <i>CD83</i>       | 0.751                      |
| brown       | <i>RPL41</i>      | 0.837                      |
| brown       | <i>RPS23</i>      | 0.827                      |
| brown       | <i>RPL10A</i>     | 0.822                      |
| brown       | <i>RPS25</i>      | 0.814                      |
| brown       | <i>RPL10</i>      | 0.808                      |
| magenta     | <i>PPP1R15A</i>   | 0.847                      |
| magenta     | <i>BHLHE40</i>    | 0.831                      |
| magenta     | <i>PLK3</i>       | 0.802                      |
| magenta     | <i>CCNL1</i>      | 0.797                      |
| magenta     | <i>CSRNP1</i>     | 0.783                      |

**Table C.5: Summary of ncMono WGCNA module genes.** Top 5 genes per module by correlation with module eigengene expression.



| Module      | Gene            | Correlation with eigengene |
|-------------|-----------------|----------------------------|
| greenyellow | <i>RPL18</i>    | 0.942                      |
| greenyellow | <i>RPL7A</i>    | 0.942                      |
| greenyellow | <i>RPS15A</i>   | 0.941                      |
| greenyellow | <i>RPS7</i>     | 0.938                      |
| greenyellow | <i>RPS23</i>    | 0.937                      |
| tan         | <i>LAG3</i>     | 0.821                      |
| tan         | <i>GZMH</i>     | 0.791                      |
| tan         | <i>MCOLN2</i>   | 0.775                      |
| tan         | <i>CLEC2D</i>   | 0.765                      |
| tan         | <i>TPRG1</i>    | 0.76                       |
| red         | <i>RAI14</i>    | 0.476                      |
| red         | <i>DEPTOR</i>   | 0.466                      |
| red         | <i>AJM1</i>     | 0.462                      |
| red         | <i>FAM171A1</i> | 0.447                      |
| red         | <i>TTC28</i>    | 0.437                      |
| turquoise   | <i>GIMAP6</i>   | 0.904                      |
| turquoise   | <i>TNFRSF1A</i> | 0.892                      |
| turquoise   | <i>GIMAP7</i>   | 0.89                       |
| turquoise   | <i>CXCR2</i>    | 0.87                       |
| turquoise   | <i>MRFAP1L1</i> | 0.862                      |
| green       | <i>CYTH1</i>    | 0.822                      |
| green       | <i>SPSB3</i>    | 0.775                      |
| green       | <i>PRSS23</i>   | 0.769                      |
| green       | <i>ANKRD12</i>  | 0.758                      |
| green       | <i>SNRK</i>     | 0.751                      |
| purple      | <i>XAF1</i>     | 0.948                      |
| purple      | <i>OAS1</i>     | 0.939                      |
| purple      | <i>IFIT3</i>    | 0.937                      |
| purple      | <i>IFI6</i>     | 0.936                      |
| purple      | <i>MX1</i>      | 0.923                      |
| brown       | <i>ZNF394</i>   | 0.881                      |
| brown       | <i>RBM39</i>    | 0.876                      |
| brown       | <i>METRNL</i>   | 0.868                      |
| brown       | <i>NR4A2</i>    | 0.862                      |
| brown       | <i>DDX6</i>     | 0.86                       |
| blue        | <i>DUT</i>      | 0.933                      |
| blue        | <i>H2AZ1</i>    | 0.93                       |
| blue        | <i>MCM7</i>     | 0.927                      |
| blue        | <i>CENPW</i>    | 0.925                      |
| blue        | <i>RRM2</i>     | 0.922                      |
| black       | <i>PSMA5</i>    | 0.904                      |
| black       | <i>PRDX3</i>    | 0.897                      |
| black       | <i>HNRNPF</i>   | 0.895                      |
| black       | <i>FARSA</i>    | 0.885                      |
| black       | <i>MRPS18C</i>  | 0.878                      |
| salmon      | <i>JCHAIN</i>   | 0.878                      |

Table C.6 continued from previous page

| Module | Gene            | Correlation with eigengene |
|--------|-----------------|----------------------------|
| salmon | <i>IGKV3-20</i> | 0.825                      |
| salmon | <i>S100A9</i>   | 0.815                      |
| salmon | <i>LYZ</i>      | 0.797                      |
| salmon | <i>IGKC</i>     | 0.786                      |

Table C.6: Summary of NK cell WGCNA module genes. Top 5 genes per module by correlation with module eigengene expression.

| Module    | Gene            | Correlation with eigengene |
|-----------|-----------------|----------------------------|
| grey      | <i>KIAA0513</i> | 0.547                      |
| grey      | <i>FUOM</i>     | 0.534                      |
| grey      | <i>EIF2AK4</i>  | 0.512                      |
| grey      | <i>ABCD4</i>    | 0.509                      |
| grey      | <i>CERCAM</i>   | 0.495                      |
| brown     | <i>SP100</i>    | 0.796                      |
| brown     | <i>IFITM1</i>   | 0.778                      |
| brown     | <i>HLA-E</i>    | 0.769                      |
| brown     | <i>OAS1</i>     | 0.768                      |
| brown     | <i>IFI35</i>    | 0.767                      |
| turquoise | <i>RAN</i>      | 0.917                      |
| turquoise | <i>NPM1</i>     | 0.907                      |
| turquoise | <i>TPI1</i>     | 0.897                      |
| turquoise | <i>SNRPD1</i>   | 0.886                      |
| turquoise | <i>XRCC5</i>    | 0.883                      |
| blue      | <i>RPL19</i>    | 0.92                       |
| blue      | <i>RPL14</i>    | 0.9                        |
| blue      | <i>RPS13</i>    | 0.9                        |
| blue      | <i>RPL37</i>    | 0.891                      |
| blue      | <i>RPL35A</i>   | 0.889                      |

Table C.7: Summary of plasmablast WGCNA module genes. Top 5 genes per module by correlation with module eigengene expression.

| Cell type | Module color | Assigned name    |
|-----------|--------------|------------------|
| B cell    | cyan         | cycling          |
| B cell    | greenyellow  | ER_stress        |
| B cell    | purple       | IFN_resp         |
| B cell    | pink         | integrin_sig     |
| B cell    | turquoise    | IRF2.ZF-C2H2     |
| B cell    | green        | mature.B.cell    |
| B cell    | midnightblue | mito_translation |
| B cell    | black        | OxPhos           |
| B cell    | blue         | p38MAPK.AP1      |
| B cell    | magenta      | ribosomal        |
| B cell    | grey         | unassigned_genes |

Table C.8 continued from previous page

| Cell type  | Module color | Assigned name    |
|------------|--------------|------------------|
| B cell     | brown        | wnt.notch        |
| CD4 T cell | lightgreen   | ambient_rna      |
| CD4 T cell | black        | cycling          |
| CD4 T cell | pink         | IFN_resp         |
| CD4 T cell | cyan         | integrin_sig     |
| CD4 T cell | turquoise    | IRF2.ZF-C2H2     |
| CD4 T cell | lightyellow  | JAK_STAT.IL_sig  |
| CD4 T cell | green        | p38MAPK.AP1      |
| CD4 T cell | midnightblue | ribosomal        |
| CD4 T cell | grey         | unassigned_genes |
| CD8 T cell | lightcyan    | ambient_rna      |
| CD8 T cell | blue         | cycling          |
| CD8 T cell | midnightblue | IFN_resp         |
| CD8 T cell | yellow       | IRF2.ZF-C2H2     |
| CD8 T cell | purple       | mito_translation |
| CD8 T cell | pink         | NK_activation    |
| CD8 T cell | black        | p38MAPK.AP1      |
| CD8 T cell | magenta      | ribosomal        |
| CD8 T cell | grey         | unassigned_genes |
| CD8 T cell | cyan         | vesicle_loc      |
| cMono      | grey60       | cycling          |
| cMono      | lightgreen   | FKBP5.CD163      |
| cMono      | black        | IFN_resp         |
| cMono      | turquoise    | IRF2.ZF-C2H2     |
| cMono      | tan          | OxPhos           |
| cMono      | lightcyan    | p38MAPK.AP1      |
| cMono      | green        | ribosomal        |
| cMono      | grey         | unassigned_genes |
| ncMono     | yellow       | cell_adhesion    |
| ncMono     | greenyellow  | cell_growth      |
| ncMono     | blue         | Cilium.PECAM1    |
| ncMono     | black        | IC_Sig           |
| ncMono     | green        | IFN_resp         |
| ncMono     | pink         | IL-1_pathway     |
| ncMono     | purple       | IRF2.ZF-C2H2     |
| ncMono     | salmon       | OxPhos           |
| ncMono     | magenta      | p38MAPK.AP1      |
| ncMono     | brown        | ribosomal        |
| ncMono     | grey         | unassigned_genes |
| NK cell    | salmon       | ambient_rna      |
| NK cell    | blue         | cycling          |
| NK cell    | red          | ECM_organisation |
| NK cell    | purple       | IFN_resp         |
| NK cell    | turquoise    | IRF2.ZF-C2H2     |
| NK cell    | green        | ITK.STAT4        |

**Table C.8 continued from previous page**

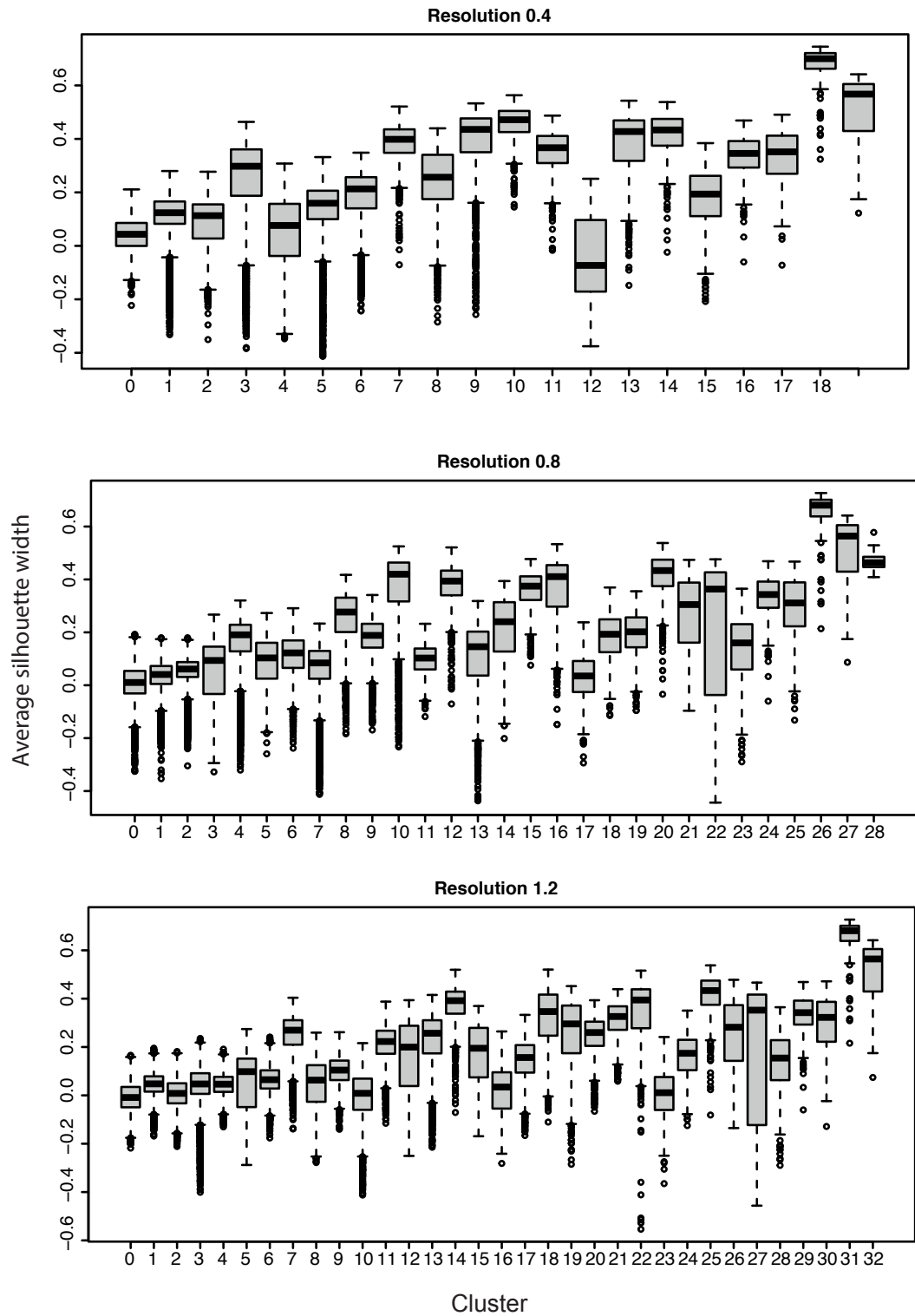
| <b>Cell type</b> | <b>Module color</b> | <b>Assigned name</b> |
|------------------|---------------------|----------------------|
| NK cell          | tan                 | leuk_prolif          |
| NK cell          | black               | mito_translation     |
| NK cell          | brown               | p38MAPK.AP1          |
| NK cell          | greenyellow         | ribosomal            |
| PB               | turquoise           | cycling              |
| PB               | brown               | IFN_resp             |
| PB               | blue                | ribosomal            |
| PB               | grey                | unassigned_genes     |

**Table C.8:** Summary of WGCNA module annotations.

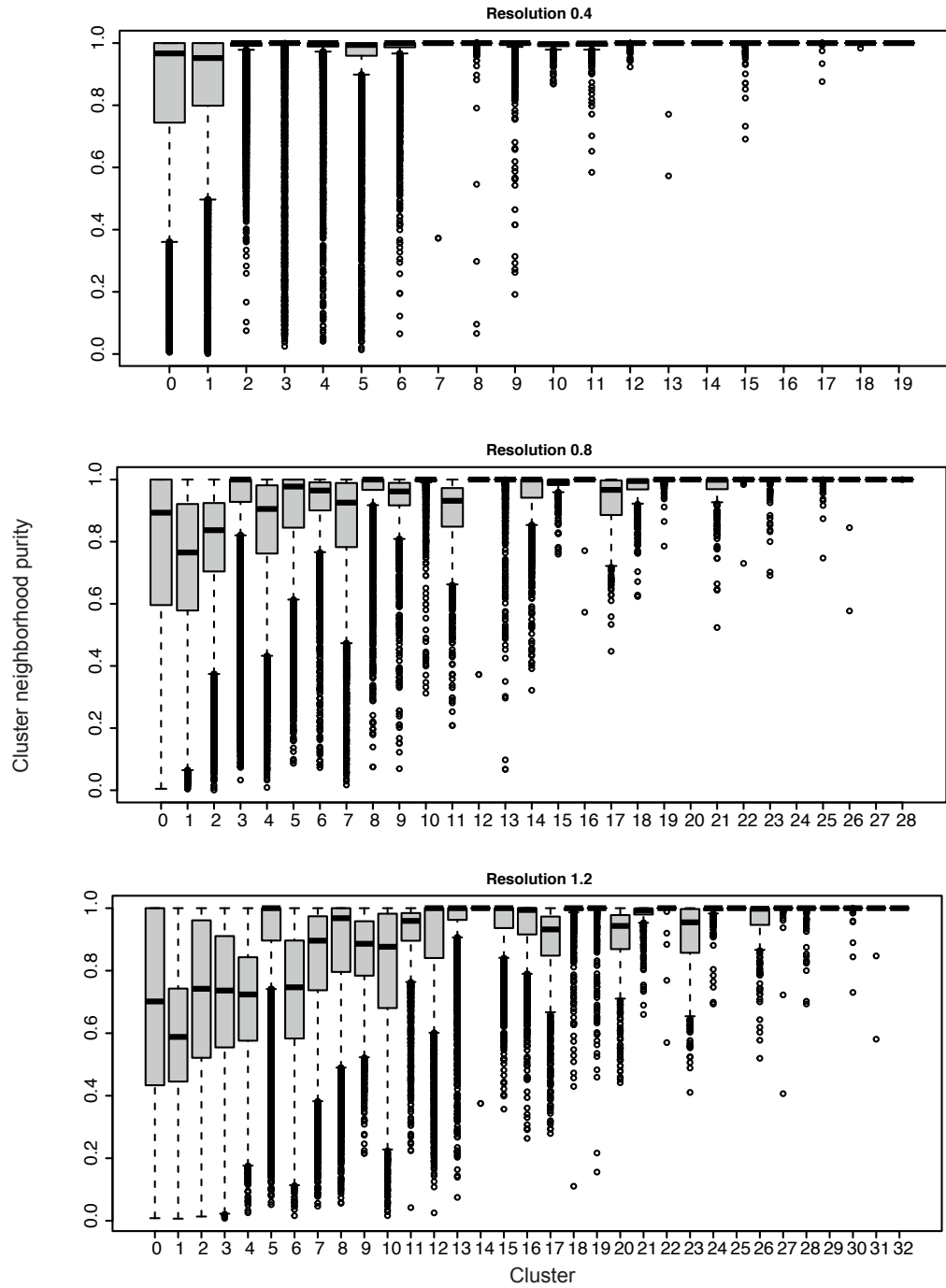
# D

## APPENDIX TO CHAPTER 5

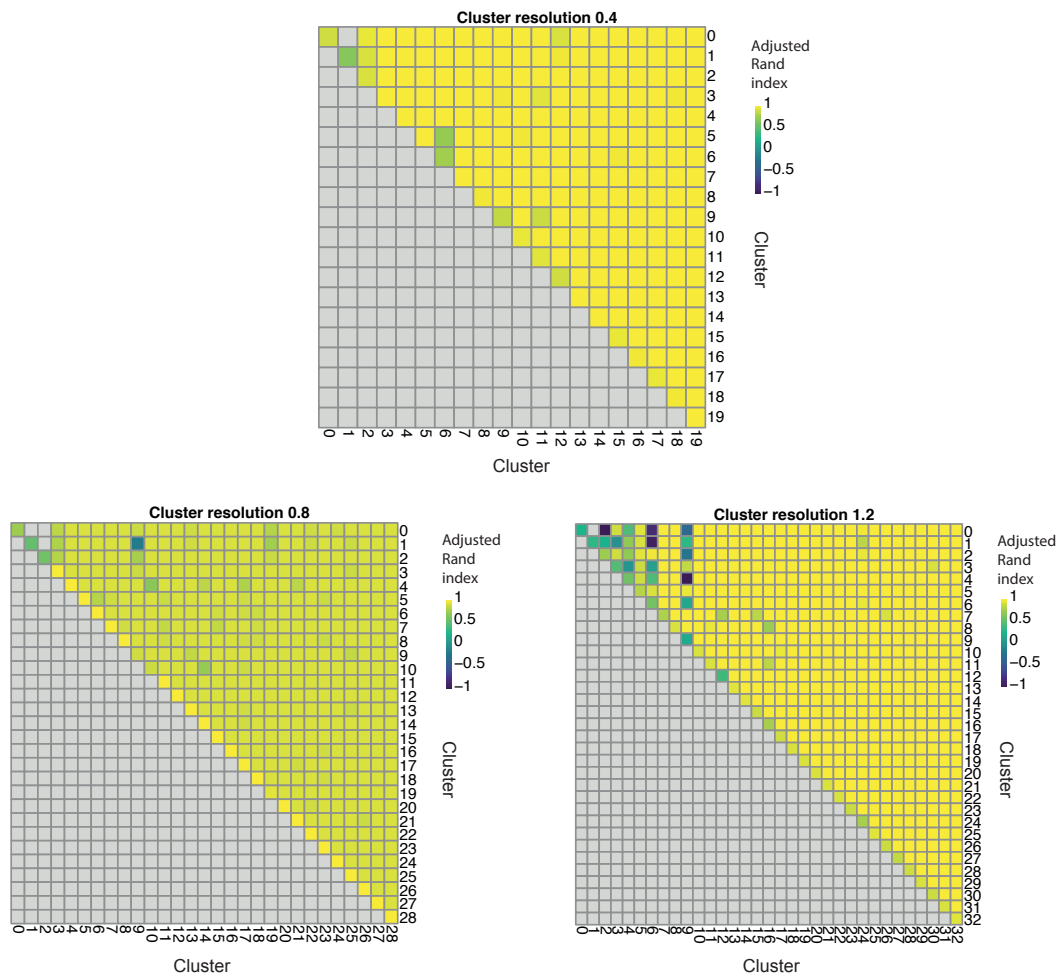
---



**Figure D.1:** scRNA-seq + Abseq clustering evaluation by average silhouette width (ASW). ASW of cell clusters at clustering resolutions of 0.4 (top), 0.8 (middle) and 1.2 (bottom).



**Figure D.2: scRNA-seq + Abseq clustering evaluation by cluster neighborhood purity.** Cluster neighborhood purity of cell clusters at clustering resolutions of 0.4 (top), 0.8 (middle) and 1.2 (bottom).



**Figure D.3: scRNA-seq + Abseq clustering evaluation by bootstrap resampling.** Cluster adjusted Rand indices for 30 iteration bootstrap resampling of cells for clustering resolutions of 0.4 (top), 0.8 (bottom left) and 1.2 (bottom right).



| Program 1 | Program 2  | Program 3 | Program 4  | Program 5 | Program 6 | Program 7 |
|-----------|------------|-----------|------------|-----------|-----------|-----------|
| OLAH      | CXCL8      | MCEMP1    | MT-ND2     | GBP5      | RSAD2     | S100A8    |
| TXNIP     | CXCR4      | IL1R2     | MT-ND1     | GBP1      | IFIT1     | S100A9    |
| BCAT1     | MNDA       | PLAC8     | MT-ND4     | GBP4      | ISG15     | CST7      |
| IL1R2     | DUSP6      | CYP1B1    | MT-CO2     | P2RY14    | IFI6      | CD177     |
| CREBRF    | BTG2       | HPGD      | MT-CYB     | APOL6     | MX1       | S100A12   |
| HMGB2     | AC012368.1 | FCER1G    | MT-ND3     | FCGR1A    | IFIT3     | GAPDH     |
| ZBTB16    | PTGS2      | C1orf162  | MT-ATP6    | STAT1     | HERC5     | PLP2      |
| FKBP5     | MPEG1      | SERPINB1  | MT-CO3     | GBP2      | IFIT2     | S100A6    |
| KLF9      | PMAIP1     | CD163     | MT-ND5     | ANKRD22   | OAS2      | CKAP4     |
| ZNF608    | CTSS       | S100A8    | MT-ND4L    | PARP14    | OAS3      | VIM       |
| SMAP2     | MALAT1     | HP        | MT-CO1     | SAMD9L    | IFI44L    | MMP9      |
| MAOA      | SGK1       | MIR223HG  | MT-ATP8    | TRIM22    | OAS1      | FCN1      |
| INHBB     | NCOA4      | CEACAM1   | MTATP6P1   | TAP1      | CMPK2     | SLC2A3    |
| MEGF9     | RABGAP1L   | IL18R1    | MT-RNR2    | IRF1      | IFI44     | ACTB      |
| MT-RNR1   | RHOB       | ANXA1     | MTND2P28   | SAMHD1    | XAF1      | PADI4     |
| IL18R1    | OGFRL1     | ALOX5AP   | MTND1P23   | EPSTI1    | MT2A      | SELL      |
| CLEC4E    | CXCL1      | PGD       | MT-RNR1    | LAP3      | LY6E      | PGLYRP1   |
| MT-RNR2   | SEC14L1    | PLIN2     | MTCO3P12   | XAF1      | EIF2AK2   | RFLNB     |
| IGKC      | CLEC7A     | MT-CO1    | AC006299.1 | CYBB      | PNPT1     | ALPL      |
| IRS2      | FGL2       | GYG1      | NKG7       | B2M       | OASL      | APMAP     |
| IRF2BP1   | HLA-B      | ARG1      | S100A11    | UBE2L6    | IFIT5     | F5        |
| ABCA1     | CYP4F3     | PCOLCE2   | LYZ        | RNF213    | DDX58     | PGD       |
| USP10     | CCR3       | SORT1     | PFN1       | IFITM3    | DDX60     | LILRA6    |
| NFE4      | TNFAIP2    | STOM      | AC040169.1 | FYB1      | EPSTI1    | S100A4    |
| CYP1B1    | P2RY10     | PTGER2    | MTND4P12   | WARS1     | USP18     | TSPO      |
| DAAM2     | CCL3L1     | FKBP5     | CFL1       | SERPING1  | ZCCHC2    | DYSF      |
| HLA-B     | CPPED1     | S100A12   | AL353726.1 | PSTPIP2   | MX2       | GPAT3     |

Table D.1 continued from previous page

| Program 1 | Program 2  | Program3 | Program 4  | Program_5 | Program_6 | Program_7 |
|-----------|------------|----------|------------|-----------|-----------|-----------|
| SLC38A1   | CCL4L2     | CEBPB    | S100A8     | XRN1      | SPATS2L   | CDC42EP3  |
| GPX1      | PECAM1     | LDHA     | TYROBP     | PML       | RNF213    | ZDHHC19   |
| MME       | RPS6KA5    | VNN1     | PNPLA2     | CD74      | TNFSF10   | STXBP2    |
| FLT3      | PHOSPHO1   | TPT1     | MT-ND6     | PARP9     | SAMD9     | PYGL      |
| MBNL1     | SETX       | SNX3     | MTCO1P12   | ETV7      | SAMD9L    | ITGB2     |
| ACSL4     | AC023157.2 | GADD45A  | CD52       | GBP3      | TMEM123   | PFN1      |
| RGS18     | BOD1L1     | SPTLC2   | MTCO2P12   | STAT2     | PARP14    | HK3       |
| NUDT5     | ZFP36L2    | S1PR1    | FLOT2      | BATF2     | IFI16     | S100P     |
| IRAK3     | C5AR1      | H2BC5    | WAS        | APOL1     | HELZ2     | MAPK14    |
| MAP3K2    | SMCHD1     | UGCG     | LINC00907  | PSMB9     | PARP12    | METTL9    |
| SPTLC2    | JARID2     | CD53     | SLC25A37   | TIFA      | IFIH1     | PKM       |
| PIAS2     | BAZ2B      | PFKFB3   | IL22RA2    | IFITM1    | IRF7      | CR1       |
| GRB10     | TOPORS     | CD55     | CSF3R      | VAMP5     | TNFSF13B  | SYNE1     |
| HLA-E     | JUND       | GNS      | LTF        | FCGR1B    | PPMIK     | ITGAM     |
| TOB1      | CHI3L1     | SLA      | AL035446.2 | GCH1      | NT5C3A    | RNF10     |
| TPST1     | HSH2D      | AGFG1    | RETN       | IFIT3     | ZBP1      | GYG1      |
| S100P     | TAGLN2     | BCL2A1   | LRP1       | ERAP2     | DHX58     | CFL1      |
| AKAP13    | FTL        | SOC3     | FLOT1      | TAP2      | BST2      | VNN2      |
| SYNE1     | PTAFR      | CLEC4D   | R3HDM4     | PLAAT4    | DDX60L    | TIMP1     |
| RIPOR2    | YPEL3      | CD44     | AC048382.2 | IFIH1     | SP100     | RAC2      |
| DHRS9     | ATP2B1     | MYL6     | IL31RA     | APOL2     | TRIM22    | GRN       |
| NUDT16    | CXCR2      | FAM20A   | CCDC183    | TRAFFD1   | TRIM14    | ADD3      |
| RPH3A     | HLA-A      | CTSB     | GRN        | NUB1      | STAT2     | ZYX       |

Table D.1: Mature neutrophil cNMF programs 1-7. Top 50 genes per program.

| Program_8 | Program_9 | Program_10 | Program_11 | Program_12 | Program_13 | Program_14 |
|-----------|-----------|------------|------------|------------|------------|------------|
| IL1RN     | FOS       | MT-CO1     | PTMA       | SLPI       | IL1B       | RGS2       |
| TNFAIP6   | EGR1      | SERPINB9   | SIGLEC10   | PI3        | NFKBIA     | PRKDC      |
| MARCKS    | DUSP1     | CD52       | KCTD12     | SIGLEC10   | ZFP36L1    | PROK2      |
| CCL3      | FOSB      | GGT5       | MTDH       | HPSE       | ADM        | ARHGEF40   |
| CCL4      | ZFP36     | STAC       | NCL        | CX3CR1     | SOC3       | BASP1      |
| TNFAIP3   | JUNB      | MTND2P28   | DYNLL1     | FTH1       | NIBAN1     | APMAP      |
| GPR84     | IER2      | PLAC8      | ANP32B     | MT-RNR1    | NAMPT      | ACSL1      |
| CLIC4     | RGS2      | ALPL       | UHMK1      | S100A4     | FFAR2      | GCA        |
| PLEK      | PTGS2     | S100A9     | GAPT       | LYZ        | BASP1      | MEGF9      |
| SLAMF7    | BTG2      | IL21R      | CANX       | AC073172.1 | NFKBIZ     | SLC25A37   |
| NFKBIA    | KLF2      | ZDHHC19    | RPL32      | HES1       | PIM2       | CHI3L1     |
| NBN       | G0S2      | AHR        | TXNIP      | MT-RNR2    | HCAR3      | H2AC6      |
| AZIN1-AS1 | MCL1      | CLU        | CD74       | CASP5      | KCNJ2      | IFITM2     |
| SOD2      | EGR3      | UBXN11     | DDX24      | THBD       | SOD2       | MALAT1     |
| CLEC4E    | KLF6      | LAIR1      | RPS24      | TMSB4X     | FCGR2A     | NAMPT      |
| C3AR1     | MYADM     | MS4A4A     | RPS28      | FGR        | ZFP36      | CXCR2      |
| GCH1      | TSC22D3   | POU2F2     | MPHOSPH8   | MARCKSL1   | IER3       | TMX4       |
| CYBB      | ZFP36L2   | SLAMF7     | CYBB       | CXCL16     | SLC25A37   | SELL       |
| BIRC3     | JUND      | SERPINI1   | FCN1       | FCER1G     | CSF2RB     | OSBPL1A    |
| CLEC4D    | PPP1R15A  | CCDC170    | SLTM       | CXCR4      | PRDM1      | MRVI1      |
| TNFAIP2   | KLF4      | PGM5       | CTSD       | MAFB       | CREB5      | TLR4       |
| ZC3H12A   | S100A4    | TSPO       | RBM25      | RHOQ       | LRG1       | LRP10      |
| CCRL2     | TAGLN2    | PALLD      | YBX1       | ADGRE2     | DSC2       | CD55       |
| SGPP2     | S100A6    | ANKRD34B   | CTSC       | PAG1       | ACSL1      | OTX1       |
| SAMSN1    | ACTB      | CD44       | SEC31A     | LTA4H      | JUNB       | SLA        |
| G0S2      | SAT1      | SH3BGRL3   | NEATC2IP   | CD9        | FCGR3B     | SORL1      |
| SLC39A8   | ZFP36L1   | TSPAN2     | RPS6       | CD14       | SLC2A3     | IFI16      |

Table D.2 continued from previous page

| Program_8 | Program_9  | Program_10 | Program_11 | Program_12 | Program_13 | Program_14 |
|-----------|------------|------------|------------|------------|------------|------------|
| GBP1      | C5AR1      | CD93       | RPS29      | ITGA4      | LITAF      | ATP2B4     |
| IRAK2     | VIM        | GRN        | U2SURP     | SERPINA1   | FTH1       | MME        |
| FTH1      | EIF1       | S100A4     | HNRNPR     | S100A6     | NAIP       | SLC11A1    |
| ANKRD22   | THBD       | XBP1       | RPS27      | BHLHE40    | LYN        | EVI2B      |
| RNF144B   | ARHGAP9    | SCPEP1     | NAP1L1     | CD84       | FYB1       | BCL6       |
| SERPINB9  | MT-RNR1    | CD300E     | DDX18      | LINC02158  | BCL2L11    | ALDH2      |
| AQP9      | ATP2B1-AS1 | MS4A6A     | PCYOX1     | FTH1P10    | CSF3R      | MGAM       |
| LIMS1     | TREM1      | MNDA       | SRP72      | IL2RG      | G0S2       | RESF1      |
| ALOX5AP   | H3-3A      | ETS2       | LYZ        | TIMP1      | IRF1       | CYSTM1     |
| TNIP1     | ACTG1      | S100A8     | PPP2R5E    | LRRC25     | KREMEN1    | IGF2R      |
| RILPL2    | ID2        | BCL2L11    | RPL34      | TXNIP      | PROK2      | SKAP2      |
| MTF1      | H3-3B      | GRINA      | M6PR       | POU2F2     | STEAP4     | LITAF      |
| ACOD1     | IQGAP2     | SH3PXD2B   | BRI3BP     | CD48       | IER5       | PTEN       |
| GRINA     | CORO1A     | UCP2       | CALM1      | BRI3       | LIMK2      | ITM2B      |
| SLC11A2   | CPPED1     | GIMAP8     | GRK3       | YBX3       | CST7       | USP10      |
| ICAM1     | OSM        | S100A12    | RPL21      | LST1       | CR1        | KIAA0319L  |
| SLC2A6    | CTSS       | BANK1      | PRPF4B     | FNIP2      | TNFAIP3    | BCL2A1     |
| ARHGAP31  | TUBA1A     | CSF3R      | KLF10      | HLA-A      | ICAM1      | CASP4      |
| TFEC      | ANXA1      | PNPLA6     | PRPF40A    | PSAP       | BCL3       | RIPOR2     |
| IVNS1ABP  | AP003733.3 | GPCPD1     | NUDT3      | EFHD2      | PTPRC      | FAM126B    |
| CKAP4     | C16orf54   | PIK3AP1    | PABPC1     | MYD88      | CCR1       | SRGN       |
| EHD1      | WLS        | ADGRE5     | SH3BGR13   | CHST7      | BIRC3      | ZNF281     |
| IL1B      | USP10      | CISH       | HLA-A      | PILRA      | CASP4      | NEAT1      |

Table D.2: Mature neutrophil cNMF programs 8-14. Top 50 genes per program.

| Program_1 | Program_2 | Program_3 | Program_4 | Program_5  | Program_6 | Program_7  | Program_8 | Program_9 |
|-----------|-----------|-----------|-----------|------------|-----------|------------|-----------|-----------|
| MMP9      | FOS       | RSAD2     | MT-CYB    | FTH1       | GBP1      | TXNIP      | PLAC8     | PROK2     |
| S100A9    | RGS2      | IFIT1     | MT-ND4    | SOD2       | GBP5      | OLAH       | MMP8      | GCA       |
| GAPDH     | DUSP1     | IFI6      | MT-ND2    | ALPL       | GBP4      | ZBTB16     | MT-CO1    | IFITM2    |
| RFLNB     | CTSS      | MX1       | MT-CO2    | CXCL16     | FCGR1A    | SMAP2      | IL1R2     | CST7      |
| S100A12   | BTG2      | ISG15     | MT-ND3    | TAGLN2     | P2RY14    | FKBP5      | CEACAM1   | ACSL1     |
| S100A8    | MME       | OAS3      | MT-ND1    | CD14       | APOL6     | IL1R2      | TPT1      | SLC25A37  |
| PKM       | PTGS2     | HERC5     | MT-ATP6   | MT-RNR1    | GBP2      | CD163      | MCEMP1    | BASP1     |
| PGLYRP1   | CYP4F3    | IFIT3     | MT-ND5    | PI3        | STAT1     | IRF2BPL    | SERPINB1  | NAMPT     |
| PADI4     | FGL2      | IFIT2     | MT-CO3    | TIMP1      | ANKRD22   | IL18R1     | LTF       | CYSTM1    |
| S100A6    | MNDA      | OAS1      | MT-ND4L   | ADGRE2     | IRF1      | MACIR      | MS4A4A    | PFKFB3    |
| FCN1      | RESF1     | IFI44     | MT-ATP8   | CCR1       | TRIM22    | SIGLEC10   | S100A12   | CD55      |
| PLBD1     | G0S2      | OAS2      | MT-CO1    | ADM        | TAP1      | DAAM2      | GYG1      | UGCG      |
| EMB       | CXCL8     | IFI44L    | MTATP6P1  | LTA4H      | SAMHD1    | KLF9       | STOM      | GYG1      |
| CTSH      | FTL       | XAF1      | FCN1      | ZFP36L1    | SAMD9L    | CYP1B1     | HP        | TNFSF13B  |
| CD177     | NCOA4     | CMPK2     | LYZ       | HPSE       | PARP14    | HLA-B      | LDHA      | ANXA3     |
| HMGB2     | ZFP36L2   | LY6E      | MTND2P28  | NIBAN1     | EPSTI1    | IRS2       | DHRS9     | BCL2A1    |
| CKAP4     | OGFRL1    | EIF2AK2   | MTND1P23  | ANKRD34B   | CYBB      | DGKH       | CYP1B1    | FLOT2     |
| VIM       | JUNB      | MT2A      | WSB1      | FCGR2A     | XAF1      | CSGALNACT2 | S100A8    | OSBPL1A   |
| BPI       | CXCR4     | IFIT5     | CD14      | AC073172.1 | WARS1     | BCAT1      | UBE2H     | LITAF     |
| LGALS1    | C5AR1     | DDX60     | NKG7      | SLPI       | CD74      | NUDT16     | VNN1      | MGAM      |
| APMAP     | FCGR3B    | ZCCHC2    | CYP1B1    | S100A4     | PLEK      | ARG1       | PFKFB2    | IL4R      |
| S100P     | KLF2      | EPSTI1    | GRN       | CREB5      | NBN       | TOB1       | PCOLCE2   | SLA       |
| SH3BGRL3  | MALAT1    | TMEM123   | CLEC12A   | BCL2L11    | GCH1      | ABCA1      | CYSTM1    | MRVI1     |
| S100A4    | ZFP36     | OASL      | SIGLEC10  | BHLHE40    | GBP3      | PAG1       | C1orf162  | F5        |
| CTSD      | JUND      | PNPT1     | HP        | ZDHHC19    | TIFA      | SPTLC2     | PLIN2     | MCEMP1    |
| ALOX5AP   | SMCHD1    | MX2       | IFI6      | SLAMF7     | UBE2L6    | MBNL1      | LAIR1     | RNF10     |
| TSPO      | CHI3L1    | SAMD9     | CFL1      | LAIR1      | SLAMF7    | TLR2       | HGF       | FRMD4B    |

Table D.3 continued from previous page

| Program_1 | Program_2  | Program_3 | Program_4 | Program_5 | Program_6 | Program_7 | Program_8 | Program_9 |
|-----------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| GRN       | YPEL3      | DDX58     | CLEC5A    | SERPINB9  | IFITM1    | GAS7      | FCER1G    | GADD45A   |
| FAM107B   | LUCAT1     | RNF213    | CD52      | GGT5      | IFITM3    | CLEC4E    | SNHG29    | H2BC21    |
| SYNE1     | DUSP6      | IFIH1     | PNPLA2    | ADGRE5    | APOL1     | RGS18     | CCR1      | CEBPB     |
| PFN1      | PHOSPHO1   | SPATS2L   | PFN1      | ITGAX     | B2M       | IRAK3     | PAG1      | SELL      |
| ITGAM     | FOSB       | SAMD9L    | TYROBP    | CLU       | SERPING1  | KLHL8     | SFXN1     | PLSCR1    |
| HK3       | FTH1       | USP18     | FAM20A    | EFHD2     | TNFAIP3   | ZNF608    | PALLD     | NIBAN1    |
| CYBB      | IGF2R      | PARP14    | EMP3      | CSF3R     | VAMP5     | ASPH      | ANXA1     | CASP4     |
| CST7      | RHOB       | IFI16     | AIF1      | TCF7L2    | BATF2     | TSC22D3   | PFKFB3    | AQP9      |
| ARG1      | SGK1       | IRF7      | MT-RNR2   | FTH1P20   | LAP3      | CREBRF    | SORT1     | C4orf3    |
| ACTB      | ITM2B      | PARP12    | RPS27     | CD93      | FCGR1B    | MAOA      | PGD       | CDKN2D    |
| ACTG1     | MPEG1      | BST2      | MCEMP1    | FYB1      | PML       | INSR      | SOC3      | BCL6      |
| CFL1      | TREM1      | PPM1K     | LAIR1     | TMSB4X    | FFAR2     | SLC38A1   | MIAT      | IFI16     |
| SNX3      | SEC14L1    | SHISA5    | CYBB      | RHOQ      | FYB1      | ITPKC     | GALNT2    | SLC2A3    |
| SLCO4C1   | B2M        | STAT2     | S100A8    | BASP1     | IFIH1     | MEGF9     | CD177     | NAIP      |
| CAP1      | BTG1       | TRIM14    | SERPINB9  | DSC2      | IFIT3     | INHBB     | GADD45A   | SORL1     |
| ADD3      | CXCL1      | HELZ2     | MIR223HG  | GRN       | NUCB1     | TRIM22    | CLEC4D    | MARCKS    |
| PLP2      | TAGAP      | SAMHD1    | ITGB2     | GAPDH     | CD274     | CD53      | CD44      | PTPRC     |
| MTARC1    | AC023157.2 | PML       | SIGLEC14  | SLC2A3    | TNFAIP6   | EVI2B     | LCN2      | PYGL      |
| CAPNS1    | CXCR2      | HERC6     | STING1    | CD82      | NFKBIA    | MAN2A2    | CLINT1    | RNF149    |
| PNPLA2    | CLEC7A     | DDX60L    | DMGDH     | SLC15A4   | GIMAP4    | ADAM17    | IL18R1    | IDI1      |
| CAPG      | ANP32A     | DHX58     | RETN      | PIK3AP1   | STAT2     | RIPOR2    | TRMT6     | MSL3      |
| CEBPA     | HSH2D      | LGALS9    | S100A12   | POU2F2    | PLAAT4    | C1orf162  | BPI       | WSB1      |
| RAC2      | BOD1L1     | ZBP1      | RPS20     | ETS2      | XRN1      | CALM2     | CD63      | SNX3      |

Table D.3: S100A8/9 high neutrophil cNMF programs. Top 50 genes per program.



| Program_1  | Program_2 | Program_3 | Program_4 | Program_5 | Program_6  | Program_7 | Program_8 |
|------------|-----------|-----------|-----------|-----------|------------|-----------|-----------|
| CD2        | MMP8      | FCGR3B    | CEACAM8   | LCN2      | IFI6       | S100A6    | IL1R2     |
| PYHIN1     | HSP90B1   | RGS2      | LYZ       | RETN      | XAF1       | EMB       | CEBPB     |
| TNFRSF25   | MT-ND4    | FTH1      | OLR1      | NKG7      | IFI44      | VIM       | GYG1      |
| MAN1C1     | MT-ND5    | ITM2B     | CAMP      | GPI       | OAS2       | S100A9    | NAMPT     |
| LCK        | CYBB      | MME       | CEACAM6   | CFL1      | MX1        | RFLNB     | SOCS3     |
| DLGAP3     | SERPINB2  | CXCR2     | MS4A3     | CYSTM1    | OAS3       | MMP9      | PLAC8     |
| AL603839.3 | HP        | NAMPT     | LTF       | S100A9    | RSAD2      | CD177     | SERPINB1  |
| RRP15      | HSPA5     | IFITM2    | BPI       | NAIP      | IFIT3      | S100A4    | CCR1      |
| LIN9       | MT-CO1    | HLA-C     | CHI3L1    | MMP8      | IFI44L     | PLBD1     | LITAF     |
| AL591848.4 | MT-CO2    | HLA-E     | RNASE2    | LDLR      | GBP1       | GAPDH     | CYP1B1    |
| DPP4       | HIPK2     | TAGLN2    | CRISP3    | PFN1      | ISG15      | LGALS1    | FYB1      |
| SPC25      | MT-ND3    | SRGN      | DEFA3     | CCND3     | SAMD9L     | CTSH      | SRGN      |
| LINC01825  | LTF       | PSAP      | PGLYRP1   | ALOX5AP   | IFIT2      | FAM107B   | VNN2      |
| CD28       | MT-ATP6   | YPEL3     | RGCC      | TSPO      | SAMHD1     | S100A12   | TRIM22    |
| ICOS       | MT-ATP8   | CTSS      | CD24      | GAPDH     | STAT1      | ARPC5     | MCEMP1    |
| WNT10A     | S100A12   | B2M       | FXN       | PIM1      | AC016542.3 | TNFAIP6   | CR1       |
| HJURP      | EEF1A1    | MT-RNR2   | ABCA13    | TFRC      | IFIT1      | CTSD      | TPT1      |
| TRAIP      | RPS24     | ZFP36L1   | LINC02009 | S100A8    | EPSTI1     | SH3BGRL3  | ACSL4     |
| CMSS1      | VNN1      | SMCHD1    | C20orf27  | SFXN1     | CMPK2      | SLC2A3    | NIBAN1    |
| HEG1       | PDIA6     | EVI2B     | CPNE3     | CHCHD7    | OAS1       | CAP1      | BCL6      |
| NSG1       | MIR223HG  | TREM1     | ITGA9     | GRINA     | LGALS9     | CAB39     | GLUL      |
| THBS4      | MT-ND1    | C5AR1     | CEBPE     | PGK1      | IFIH1      | GCA       | LAIR1     |
| KIF20A     | PGAM1     | IGF2R     | SPN       | CST7      | RAD51AP1   | CCDC69    | CDK5RAP2  |
| PROB1      | MT-ND2    | BTG1      | SLC43A1   | UGCG      | TRIM22     | GABARAPL2 | IFITM1    |
| AL139393.3 | TPT1      | RESF1     | DEFA4     | LDHA      | APOL6      | TMSB4X    | BASP1     |
| OSBPL3     | PLBD1     | SOD2      | SLPI      | RGL4      | GBP4       | APMAP     | FCER1G    |
| AUTS2      | ACTG1     | APOBEC3A  | TCN1      | ENO1      | AC139495.1 | SLC40A1   | IL18R1    |

Table D.4 continued from previous page

| Program_1  | Program_2 | Program_3 | Program_4  | Program_5 | Program_6  | Program_7 | Program_8 |
|------------|-----------|-----------|------------|-----------|------------|-----------|-----------|
| AC105052.4 | PABPC1    | MCL1      | SEMA6A     | RPS28     | GVQW3      | BPI       | LRG1      |
| TRBC2      | MANF      | CRISPLD2  | STOX2      | TPI1      | IRF1       | TSPAN2    | ZFP36     |
| LINC00892  | PRUNE2    | DUSP1     | CEP70      | ACTG1     | AP003071.1 | S100A8    | SLC25A37  |
| DMTN       | DHRS9     | SAT1      | AC016773.2 | KBTBD7    | NCKIPSD    | SNX3      | PLIN2     |
| FZD3       | CORO2A    | H3-3A     | CDK20      | EMP3      | MT-CO2     | CAST      | ANKRD34B  |
| AC022730.4 | OIP5-AS1  | HLA-B     | FLT1       | CNN2      | TAP1       | PGLYRP1   | CD53      |
| LINC00092  | TMSB4X    | CMTM2     | AC012181.2 | IDI1      | PARP14     | HMGB2     | TLR4      |
| NPDC1      | LINC01806 | RBP7      | CDC14B     | SYNGR2    | ZDHHC19    | CYP4F3    | SORL1     |
| CHEK1      | TMEM255A  | FGL2      | SERPINB10  | LTF       | BTBD11     | LIMS1     | FCGR2A    |
| FGFBP3     | HACD3     | ANP32A    | SLC39A11   | OLFM4     | MT-CYB     | ITGB2     | HPGD      |
| CCDC65     | ABCC6     | MALAT1    | AZU1       | LILRB4    | AC133644.2 | MNDA      | FKBP5     |
| SLC7A1     | RPL31     | FTH1P20   | SPNS3      | FKBP1A    | C3orf14    | TNIP1     | CTSB      |
| TRAV16     | RPS12     | JAML      | NPM1P39    | MYL6      | H4C11      | PADI4     | TNFSF13B  |
| LINC02325  | RPL12     | LSP1      | TNN        | SLC16A3   | EHD3       | ARL6IP6   | PCOLCE2   |
| CEP170B    | COX7A2L   | IFITM3    | EIF2AK3-DT | CD24      | GBP2       | BRI3      | SH3GLB1   |
| APBA2      | MT-CYB    | JUNB      | EPM2A      | RPLP2     | GBP5       | MYADM     | H3-3B     |
| BUB1B      | YBX1      | PILRA     | DPY19L1P1  | PNPLA6    | MT-ND4     | SLC39A8   | UBE2H     |
| AC025279.1 | CANX      | HLA-A     | RPL7P32    | RPS2      | FCN1       | CALM1     | SELL      |
| KIF18B     | BPI       | ADGRE3    | MS4A2      | DHRS9     | DDX60      | ANXA3     | PGD       |
| SKAP1      | SVIP      | LST1      | AL731569.1 | SORT1     | MT-ATP6    | ACTR2     | GPR137B   |
| HOXB4      | MT-ND4L   | ABTB1     | RPS27P21   | DHCR7     | AL451085.3 | MYL12A    | SERPINA1  |
| AC022211.3 | RPL37A    | STEAP4    | RPL7P46    | CAPG      | AC022126.1 | ORAI2     | FOS       |
| TSEN54     | ACTB      | FTL       | AC130689.1 | CD163     | BX842570.1 | GPAT3     | PRKAR1A   |

Table D.4: *IL1R2*+ immature neutrophil cNMF programs. Top 50 genes per program.



| Program_1 | Program_2 | Program_3 | Program_4 | Program_5  | Program_6  | Program_7 | Program_8 |
|-----------|-----------|-----------|-----------|------------|------------|-----------|-----------|
| BPI       | RNASE2    | SELL      | IFIT3     | CD28       | CENPF      | S100A8    | LTF       |
| HP        | DEFA3     | MMP9      | OAS3      | TRBC2      | HMG2       | CD177     | CAMP      |
| MT-RNR1   | RNASE3    | EVI2B     | OAS1      | TRAC       | PTTG1      | MMP8      | PGLYRP1   |
| H2AJ      | SLPI      | FPR1      | OAS2      | TCF7       | HSP90B1    | PLAC8     | CHIT1     |
| MS4A7     | MS4A3     | HMGB2     | MTCO2P33  | GIMAP7     | MKI67      | SERPINA1  | PADI4     |
| OLR1      | CEACAM6   | VNN1      | LY6E      | LEF1       | PTMA       | CST7      | CPNE3     |
| FSTL3     | DEFA4     | S100A12   | IFI6      | AC241585.2 | PRR11      | MCEMP1    | CRISP3    |
| SMARCD1   | SLC2A5    | VNN2      | CCR1      | LY9        | CEP55      | UGCG      | CEACAM8   |
| TPT1      | AZU1      | S100A4    | IFIT1     | SLC41A1    | PTMS       | MT-CO3    | CYBB      |
| ACP3      | BPI       | C5AR1     | EPSTI1    | TRABD2A    | TOP2A      | PFKFB3    | CCDC125   |
| EMP3      | SPN       | SRGN      | XAF1      | LINC01825  | GGH        | MT-ND1    | ACTB      |
| XIST      | TPP1      | MEGF9     | NAMPT     | ABI2       | GTSE1      | CD55      | CEACAM1   |
| CRISP3    | ALDH2     | RG2       | ZFP36L1   | MCUB       | LINC01355  | ANXA3     | LCN2      |
| APOBEC3H  | C20orf27  | S100A11   | GBP4      | ACSL6      | FDPSP1     | GYG1      | LGALS     |
| LYZ       | CD63      | H3-3A     | RSAD2     | KLHL3      | DNAH6      | PIM1      | OLFM4     |
| ITGB2-AS1 | MT-CO2    | CTSS      | IFITM1    | ADTRP      | AC010245.2 | GCA       | CPNE2     |
| S100A9    | NKG7      | S100A6    | HERC5     | FAM184A    | HCG25      | FCER1G    | S100A9    |
| GGTA1P    | MT-CO3    | SDCBP     | TRIM22    | CARD11     | MDFIC      | FYB1      | MRTFA     |
| CD52      | TMSB4X    | MXD1      | IFITM3    | RAB39B     | CBX1P1     | CAP1      | SEC63     |
| JUN       | MT-ATP6   | LIMD2     | PARP14    | MICU3      | IDO1       | SLC2A3    | FCN1      |
| LCN2      | MT-ND3    | CLEC4E    | SERPING1  | MTUS1      | AL590787.1 | CEACAM1   | LTA4H     |
| SLC16A3   | MT-CO1    | STK17B    | ALPL      | PKIA       | WFIKKN1    | TCN1      | DDX3Y     |
| TPM4      | MT-ND4    | USP10     | FFAR2     | AC068700.1 | KRT38      | STOM      | CORO1A    |
| SERPINB10 | GRN       | DUSP1     | SLAMF7    | DNLZ       | CIT        | LDHA      | SLC31A1   |
| H2BC8     | RAB13     | LILRA6    | MT2A      | MAML2      | HMGB3      | CYSTM1    | PFN1      |
| RHOB      | AIF1      | LRRK2     | NIBAN1    | KLRB1      | CENPE      | YWHAH     | ABCA13    |
| PGLYRP1   | CD24      | GAPT      | IFI44L    | NELL2      | NUCKS1     | MS4A4A    | ORM1      |

Table D.5 continued from previous page

| Program_1  | Program_2  | Program_3 | Program_4  | Program_5  | Program_6  | Program_7  | Program_8  |
|------------|------------|-----------|------------|------------|------------|------------|------------|
| AC079760.2 | MT-ND1     | FCGR3B    | KLHL5      | CCDC65     | KIF14      | PGD        | AP3S1      |
| ITM2C      | MPO        | CRISPLD2  | AQP9       | BICDL1     | H2AZ2      | DHRS9      | HVCN1      |
| LGALS1     | MT-CYB     | MAP4K4    | MX2        | GPR183     | CKAP2      | JAK3       | ADSS2      |
| MYADM      | CHI3L1     | CR1       | SECTM1     | MAF        | ASPM       | P2RY13     | PAXBP1     |
| APOBEC3C   | RPLP1      | GNG2      | KREMEN1    | MMD        | MIR223HG   | IDI1       | FKBP15     |
| MT-CO1     | RETN       | SLC2A3    | IFIT2      | TSHZ2      | HDGF       | ATP6V1C1   | CITED4     |
| TSPO       | LYZ        | GLIPR1    | APOL6      | ZHX3       | STMN1      | ACTB       | CLC        |
| SPACA6     | GSTP1      | IQGAP1    | FTH1       | ARHGEF19   | CDC25B     | RNU5F-1    | AC008771.1 |
| OGDH       | PPIA       | MGAM      | FCGR3B     | CD247      | DLGAP5     | MORN2      | TMEM63A    |
| LILRA6     | NPC2       | CMTM2     | CMPK2      | DCBLD2     | BIRC5      | LINC02569  | ARFGEF1    |
| SERPINB2   | RPS20      | PSAP      | IFI44      | TSPAN5     | SMC4       | AL731566.3 | SRRM2      |
| SVIP       | MT-ND2     | H2AC6     | SAT1       | FAM169A    | SPC24      | TBX21      | LGALS3     |
| RAC2       | VAMP8      | PTPRC     | IFITM2     | MARCHF3    | GPSM2      | MT-ND2     | NSD3       |
| TAGLN2     | AL034417.5 | TBL1X     | ADGRE2     | AL135905.1 | CENPH      | DYTN       | BLM        |
| CD177      | PTMA       | IFITM2    | ISG15      | THEMIS     | KIF15      | LRG1       | OLR1       |
| AC090152.1 | COX8A      | CPPED1    | SLC25A37   | CLK2P1     | AC013400.1 | LDLR       | CDCA7L     |
| TSPAN2     | RPS19      | TMX4      | CLEC7A     | DENND11    | RGPD4      | RGL4       | NOL8       |
| CD300C     | MED8       | CREB5     | CEBPB      | KCNH2      | AC092053.5 | HP         | DIP2A      |
| AHCYL2     | CANX       | NCF1C     | AC011346.1 | PRAG1      | AC109992.1 | PADI2      | VEGFA      |
| JUNB       | PNP        | FRAT2     | ABCC2      | SORBS3     | AL133467.1 | RTN4       | MT-ATP8    |
| ERAP2      | MT-ND5     | IL1R2     | FCGR2A     | E2F5       | LSM5       | FAR2       | CS         |
| CTSD       | NUCB2      | BCL6      | FAM20A     | LAPTM4B    | ANLN       | GRM2       | PTGES      |
| NTNG2      | SLC28A3    | RIPOR2    | LRP12      | MTSS1      | CKAP2L     | AC084864.2 | TECR       |

Table D.5: *PADI4*+ immature neutrophil cNMF programs. Top 50 genes per program.

| Program_1 | Program_2  | Program_3  | Program_4 | Program_5 | Program_6  |
|-----------|------------|------------|-----------|-----------|------------|
| KPNA2     | CTSG       | CD177      | PCNA      | FTH1      | DEFA3      |
| CCNB1     | PRSS57     | S100A12    | DUT       | ZFP36L1   | DEFA4      |
| PLK1      | CST7       | CKAP4      | DHFR      | NAMPT     | MALAT1     |
| UBE2C     | TFPI       | MMP9       | MCM3      | FCGR3B    | SRGN       |
| AURKA     | C2orf88    | CD14       | TYMS      | ALPL      | TENT5A     |
| TPX2      | SELENOP    | CD52       | GIN52     | IFITM2    | SPN        |
| ARL6IP1   | RPSAP61    | PLBD1      | HMGAI     | VNN2      | NEAT1      |
| TOP2A     | RTL8B      | FPR1       | MCM4      | SELL      | DEFT1P     |
| HMMR      | CA8        | HP         | MYBL2     | CXCR2     | ANKRD18A   |
| DLGAP5    | AP001453.5 | MMP8       | TK1       | SLC11A1   | ALAS1      |
| CENPA     | ASB2       | PADI4      | MCM7      | KREMEN1   | PNP        |
| HMGB2     | TMEM121    | LGALS1     | POLD2     | SOD2      | PRRT4      |
| CKS2      | AL008727.1 | CD53       | MCM6      | CMTM2     | P2RX1      |
| CALM2     | SLC24A3    | CLEC4E     | PTMA      | RGS2      | SERPINB1   |
| CDC20     | CYTL1      | BCL2A1     | MCM5      | FFAR2     | CELF2      |
| TTK       | CALR       | CD44       | GSTP1     | AQP9      | AC006453.3 |
| NEK2      | DEFB1      | H2AJ       | CLSPN     | FTH1P20   | AC073352.1 |
| TROAP     | PRTN3      | CTSH       | BRI3BP    | PROK2     | METTL14-DT |
| PTTG1     | SMIM24     | CRISP3     | NASP      | GBP2      | TCAF2      |
| CKS1B     | TNFRSF25   | GPR65      | TPI1      | LILRB3    | AC007541.1 |
| ANLN      | PYHIN1     | S100A6     | HSPA8     | ADM       | TNFRSF13B  |
| CEP55     | BIN1       | C5AR1      | FADS1     | RIPOR2    | CNTNAP1    |
| CENPE     | TRIM71     | COLQ       | CENPU     | IL1B      | AP000238.1 |
| CDCA8     | C3orf18    | UBE2NL     | GCHFR     | FTH1P10   | SLPI       |
| CDC25B    | GPRIN3     | MTCO2P11   | PCLAF     | FPR1      | APEX1      |
| RAD21     | BACH2      | RPL21P116  | H2AZ1     | JAML      | GRK2       |
| ASPM      | TMEM67     | AC015660.4 | CENPB     | LILRA6    | MT-RNR2    |

Table D.6 continued from previous page

| Program_1 | Program_2   | Program_3  | Program_4 | Program_5 | Program_6  |
|-----------|-------------|------------|-----------|-----------|------------|
| CDKN3     | STX17-AS1   | BMX        | CDCA7     | FCGR2A    | TACSTD2    |
| TUBA1C    | CD6         | KIAA0513   | RRM2      | LINC01948 | DEFA8P     |
| PRR11     | DIP2C       | ECE1       | TAGLN2    | SLC25A37  | ACSM3      |
| G2E3      | LOXL4       | CLEC4D     | RMI2      | CR1       | EREG       |
| CCNG2     | TC2N        | RIPOR2     | SVIP      | DUSP1     | DUSP6      |
| CDK1      | CORO7-PAM16 | TMCC2      | ASF1B     | CLEC7A    | PTPN7      |
| GPSM2     | CCL23       | PLXNA1     | TCF19     | TMCC3     | PSMB7      |
| KIF20B    | CBX2        | TPRG1      | DEK       | SMAP2     | MTND2P28   |
| KIF14     | CD226       | DNAJC5B    | CHAF1A    | CRISPLD2  | C8B        |
| NUF2      | FCGBP       | LINC02453  | UNG       | TLR2      | MTRNR2L8   |
| CENPF     | ADM5        | AC007998.4 | POGLUT3   | IFIT2     | EFNB2      |
| DEPDC1    | LINC01422   | NLRP12     | MCM2      | ITM2B     | AC012181.2 |
| SPDL1     | ELANE       | SLC2A3     | HELLS     | CYP27A1   | IDI2-AS1   |
| CKAP5     | IFT140      | CR1        | SLBP      | IFITM1    | AGAP6      |
| ARHGAP11A | SMPDL3B     | CCDC140    | ZNF367    | ZDHHHC18  | AC104763.1 |
| PIF1      | CPXM1       | LILRB3     | SLFN13    | S100A4    | CORO2A     |
| CCNB2     | STAR        | MARCKSL1   | RAD51     | C5AR1     | AKAP9      |
| KIF18A    | RPL26P30    | AP000547.3 | WDR76     | IL1RN     | MT-ATP6    |
| GAS2L3    | ZBED3       | DHRS9      | CHEK1     | LYN       | CSPP1      |
| PRC1      | SLC41A1     | LTF        | HACD3     | LIMD2     | CANX       |
| NCAPG     | AZU1        | LILRA1     | BPI       | ITGAX     | TMEM87A    |
| CAMP      | COL24A1     | LILRA5     | ITM2C     | IFITM3    | THBS4      |
| KIF11     | AUTS2       | GNG2       | UHRF1     | NIBAN1    | VAMP8      |

Table D.6: MPO+ immature neutrophil/progenitor neutrophil cNMF programs 1-6. Top 50 genes per program.

| Program_7 | Program_8 | Program_9  | Program_10 | Program_11 | Program_12 |
|-----------|-----------|------------|------------|------------|------------|
| HLA-DPB1  | S100A9    | BTF3P10    | TUBB       | S100A8     | RPS4X      |
| HLA-DPA1  | RETN      | AC004552.1 | TUBA1B     | CAP1       | EEF1A1     |
| CYP2U1    | PGLYRP1   | GLUD2      | H4C3       | FYB1       | RPS6       |
| FGD2      | BTG2      | AF274853.1 | STMN1      | TMSB4X     | SNHG29     |
| HLA-DRA   | CHIT1     | DNAJC19P9  | H2AZ1      | P2RY13     | RPL31      |
| PGBD5     | JUNB      | ATL1       | HMGB1      | CYBB       | RPS19      |
| NID1      | LCN2      | THBS1      | RRM2       | TCN1       | RPL27A     |
| ADD2      | FCN1      | AC010332.3 | TYMS       | SERPINA1   | RPS24      |
| CLEC4F    | TSPO      | GDF11      | HMG2       | MIR223HG   | RPS28      |
| FMNL2     | CEBPE     | MYC        | AURKB      | METTL9     | RPL19      |
| ARHGAP31  | CORO1A    | HSPA5      | PTMA       | STOM       | RPL34      |
| GUCY1B1   | OAZ1      | HSP90B1    | FAM111A    | LDHA       | RPS8       |
| TRIM36    | HLA-A     | CD69       | TUBA4A     | ACTB       | RPS11      |
| HAVCR2    | EMP3      | HSPH1      | SPC24      | MMP8       | RPS13      |
| HLA-DQA2  | IER2      | AC002470.1 | RRM1       | CD63       | RPS23      |
| Z95118.2  | SLC16A3   | FKBP4      | KIF11      | GCA        | RPS14      |
| SASH1     | MSRB1     | LINC01772  | MKI67      | CEACAM1    | RPS3A      |
| CLIP2     | VSIR      | AC006210.2 | TOP2A      | CPNE3      | RPL13A     |
| ARHGAP6   | NCF1C     | PLAAT4     | KIFC1      | MNDA       | RPS12      |
| RTL5      | PRAM1     | TTLL11     | PCLAF      | ALOX5AP    | RPL30      |
| FTH1P8    | HLA-C     | AL691403.2 | SPC25      | ANXA3      | RPL5       |
| SLC18A1   | KLF2      | ZNF391     | ASF1B      | CEACAM8    | RPS27      |
| GFRA2     | PNPLA2    | HSP90AA1   | PRC1       | PADI2      | RPL32      |
| SORBS3    | ACTB      | AL353708.3 | GNAS       | ARG1       | RPL37      |
| ZNF703    | CHI3L1    | PDIA6      | H2AZ2      | MCEMP1     | RPL12      |
| LYNX1     | BST2      | ODC1       | CDCA5      | B4GALT5    | RPS5       |
| PTENP1-AS | WAS       | SLC4A7     | FBXO5      | MSN        | RPL13      |

Table D.7 continued from previous page

| Program_7   | Program_8 | Program_9  | Program_10 | Program_11 | Program_12 |
|-------------|-----------|------------|------------|------------|------------|
| PTGDR2      | GNB2      | PDIA3      | KIF18B     | PFKFB3     | RPS2       |
| GRASP       | CAPZB     | MANF       | EZH2       | GLRX       | RPL10      |
| NR4A1       | RHOB      | SRD5A3     | TMPO       | BASP1      | RPL29      |
| CYSLTR2     | GRN       | AC016027.1 | H2AX       | RTN4       | RPL6       |
| SAMD4A      | EGR1      | SIGLEC12   | HMGB2      | GABARAPL2  | RPL4       |
| CKB         | EHD1      | CHAC2      | FANCI      | FOLR3      | RPS3       |
| VMO1        | CFL1      | HSPA8      | CCDC34     | ANXA1      | RPL22      |
| BCAS4       | ACAP1     | AC113410.5 | CENPK      | CALM1      | RPL37A     |
| ADGRE4P     | GNAI2     | HSP90AB1   | NUSAP1     | LITAF      | RACK1      |
| DNASE2      | DNM2      | MYH3       | IQGAP3     | IKZF1      | RPL3       |
| LINC02432   | ACTG1     | MUC19      | H3-3A      | ACTG1      | TPT1       |
| MARCHF1     | SUN2      | AC010435.1 | SMC2       | PLSCR1     | RPL15      |
| FCGR3A      | CYBC1     | DDX21      | DUT        | CHI3L1     | RPS18      |
| CD74        | ZYX       | BEX3       | FOXM1      | RAB31      | RPLP1      |
| CD300E      | IL16      | AC108022.2 | ANP32B     | OLFM4      | RPL11      |
| TLR7        | CDA       | AC073140.3 | UBE2T      | PPP1R18    | RPS21      |
| VASH1       | SEPTIN9   | AC078983.1 | E2F7       | UGCG       | RPS27A     |
| MAFB        | PTMS      | RPL13AP3   | BIRC5      | FCER1G     | RPL39      |
| ENDOD1      | LYZ       | AC009268.2 | KNL1       | GYG1       | BTF3       |
| HLA-DQA1    | OGDH      | NPIPA3     | OIP5       | CTSS       | RPS15A     |
| CX3CR1      | FBXW5     | FKBP7      | GTSE1      | IRAK3      | RPL24      |
| FAM198B-AS1 | SPI1      | GAREM2     | TMEM106C   | CD55       | RPS25      |
| AKR1B1      | CYBB      | HSPD1      | NDC80      | APMAP      | RPS29      |

Table D.7: MPO+ immature neutrophil/progenitor neutrophil cNMF programs 7-12. Top 50 genes per program.

## BIBLIOGRAPHY

---

- Abani, O., Abbas, A., Abbas, F., et al. (2021). "Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial". *The Lancet* 397: pp. 1637–1645.
- Abraham, E., Arcaroli, J., and Shenkar, R. (2001). "No Title". 166: pp. 522–530.
- Ackermann, M., Verleden, S. E., Kuehnel, M., et al. (2020). "Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19". *The New England journal of medicine* 383: pp. 120–128.
- Adams, N. M., Grassmann, S., and Sun, J. C. (2020). "Clonal expansion of innate and adaptive lymphocytes". *Nature Reviews Immunology* 20:11 20: pp. 694–707.
- Agashe, C., Chiang, D., Grishin, A., et al. (2017). "Impact of granulocyte contamination on PBMC integrity of shipped blood samples: Implications for multi-center studies monitoring regulatory T cells". *Journal of immunological methods* 449: p. 23.
- Alshetaiwi, H., Pervolarakis, N., McIntyre, L. L., et al. (2020). "Defining the emergence of myeloid-derived suppressor cells in breast cancer using single-cell transcriptomics". *Science immunology* 5.
- Amodio, M., Dijk, D. van, Srinivasan, K., et al. (2019). "Exploring single-cell data with deep multitasking neural networks". *Nature Methods* 2019 16:11 16: pp. 1139–1145.
- Anderson, N. L. and Anderson, N. G. (2002). "The human plasma proteome: history, character, and diagnostic prospects". *Molecular cellular proteomics : MCP* 1: pp. 845–867.
- Andreis, D. T. and Singer, M. (2016). "Catecholamines for inflammatory shock: a Jekyll-and-Hyde conundrum". *Intensive Care Medicine* 2016 42:9 42: pp. 1387–1397.
- Aneja, R. K., Tsung, A., Sjodin, H., et al. (2008). "Preconditioning with high mobility group box 1 (HMGB1) induces lipopolysaccharide (LPS) tolerance". *Journal of Leukocyte Biology* 84: pp. 1326–1334.
- Antcliffe, D. B., Burnham, K. L., Al-Beidh, F., et al. (2019). "Transcriptomic signatures in sepsis and a differential response to steroids from the VaNISH randomized trial". *American Journal of Respiratory and Critical Care Medicine* 199: pp. 980–986.
- Aouba, A., Baldolli, A., Geffray, L., et al. (2020). "Targeting the inflammatory cascade with anakinra in moderate to severe COVID-19 pneumonia: case series". *Annals of the rheumatic diseases* 79: pp. 1381–1382.
- Argelaguet, R., Arnol, D., Bredikhin, D., et al. (2020). "MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data". *Genome Biology* 21: pp. 1–17.
- Aschenbrenner, A. C., Mouktaroudi, M., Krämer, B., et al. (2021). "Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients". *Genome medicine* 13: p. 7.
- Ashuach, T., Reidenbach, D. A., Gayoso, A., et al. (2022). "PeakVI: A deep generative model for single-cell chromatin accessibility analysis". *Cell Reports Methods* 2: p. 100182.



- Atterton, B., Paulino, M. C., Povoas, P., et al. (2020). "Sepsis Associated Delirium". *Medicina (Kaunas, Lithuania)* 56.
- Austermann, J., Friesenhagen, J., Fassl, S. K., et al. (2014). "Alarmins MRP8 and MRP14 Induce Stress Tolerance in Phagocytes under Sterile Inflammatory Conditions". *Cell Reports* 9: pp. 2112–2123.
- Baghela, A., Pena, O. M., Lee, A. H., et al. (2022). "Predicting sepsis severity at first clinical presentation: The role of endotypes and mechanistic signatures". *eBioMedicine* 75.
- Banchereau, R., Hong, S., Cantarel, B., et al. (2016). "Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients". *Cell* 165: pp. 551–565.
- Barnaby, D., Ferrick, K., Kaplan, D. T., et al. (2002). "Heart rate variability in emergency department patients with sepsis". *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine* 9: pp. 661–670.
- Bastard, P., Rosen, L. B., Zhang, Q., et al. (2020). "Autoantibodies against type I IFNs in patients with life-threatening COVID-19". *Science* 370.
- Bauer, M., Gerlach, H., Vogelmann, T., et al. (2020). "Mortality in sepsis and septic shock in Europe, North America and Australia between 2009 and 2019-results from a systematic review and meta-analysis". *Critical Care* 24: pp. 1–9.
- Bayik, D., Zhou, Y., Park, C., et al. (2020). "Myeloid-Derived Suppressor Cell Subsets Drive Glioblastoma Growth in a Sex-Specific Manner". *Cancer discovery* 10: pp. 1210–1225.
- Bendall, S. C., Simonds, E. F., Qiu, P., et al. (2011). "Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum". *Science (New York, N.Y.)* 332: pp. 687–696.
- Bergen, V., Lange, M., Peidli, S., et al. (2020). "Generalizing RNA velocity to transient cell states through dynamical modeling". *Nature Biotechnology* 2020 38:12 38: pp. 1408–1414.
- Bernardes, J. P., Mishra, N., Tran, F., et al. (2020). "Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19". *Immunity* 53: 1296–1314.e9.
- Beutler, B. and Rietschel, E. T. (2003). "Innate immune sensing and its roots: the story of endotoxin". *Nature Reviews Immunology* 2003 3:2 3: pp. 169–176.
- Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W.-C., et al. (2020). "Graphical Abstract Highlights SARS-CoV-2 infection induces low IFN-I and-III levels with a moderate ISG response".
- Blank, C. U., Haining, W. N., Held, W., et al. (2019). "Defining 'T cell exhaustion'". *Nature Reviews Immunology* 2019 19:11 19: pp. 665–674.
- Bo, L., Wang, F., Zhu, J., et al. (2011). "Granulocyte-colony stimulating factor (G-CSF) and granulocyte-macrophage colony stimulating factor (GM-CSF) for sepsis: a meta-analysis". *Critical care (London, England)* 15.
- Boettcher, S., Gerosa, R. C., Radpour, R., et al. (2014). "Endothelial cells translate pathogen signals into G-CSF-driven emergency granulopoiesis". *Blood* 124: pp. 1393–1403.
- Boisrame-Helms, J., Kremer, H., Schini-Kerth, V., et al. (2013). "Endothelial Dysfunction in Sepsis". *Current Vascular Pharmacology* 11: pp. 150–160.



- Bone, R. C., Balk, R. A., Cerra, F. B., et al. (1992). "Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine". *Chest* 101: pp. 1644–1655.
- Boomer, J. S., To, K., Chang, K. C., et al. (2011). "Immunosuppression in Patients Who Die of Sepsis and Multiple Organ Failure". *JAMA* 306: pp. 2594–2605.
- Borovikova, L. V., Ivanova, S., Zhang, M., et al. (2000). "Vagus nerve stimulation attenuates the systemic inflammatory response to endotoxin". *Nature* 405: pp. 458–462.
- Brady, G., Barbara, M., and Iscove, N. N. (1990). "Representative in Vitro cDNA Amplification From Individual Hemopoietic Cells and Colonies". *Methods in molecular and cellular biology* 2: pp. 17–25.
- Braudeau, C., Salabert-Le Guen, N., Chevreuil, J., et al. (2021). "An easy and reliable whole blood freezing method for flow cytometry immunophenotyping and functional analyses". *Cytometry Part B: Clinical Cytometry* 100: pp. 652–665.
- Brealey, D., Brand, M., Hargreaves, I., et al. (2002). "Association between mitochondrial dysfunction and severity and outcome of septic shock". *Lancet (London, England)* 360: pp. 219–223.
- Brenner, S. (2010). "Sequences and consequences". *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: pp. 207–212.
- Bronte, V., Brandau, S., Chen, S. H., et al. (2016). "Recommendations for myeloid-derived suppressor cell nomenclature and characterization standards". *Nature Communications* 2016 7:1 7: pp. 1–10.
- Brudecki, L., Ferguson, D. A., McCall, C. E., et al. (2012). "Myeloid-Derived Suppressor Cells Evolve during Sepsis and Can Enhance or Attenuate the Systemic Inflammatory Response". *Infection and Immunity* 80: p. 2026.
- Brunet, J. P., Tamayo, P., Golub, T. R., et al. (2004). "Metagenes and molecular pattern discovery using matrix factorization". *Proceedings of the National Academy of Sciences of the United States of America* 101: pp. 4164–4169.
- Burgener, S. S. and Schroder, K. (2020). "Neutrophil Extracellular Traps in Host Defense". *Cold Spring Harbor perspectives in biology* 12: pp. 1–15.
- Burn, E., Tebé, C., Fernandez-Bertolin, S., et al. (2021). "The natural history of symptomatic COVID-19 during the first wave in Catalonia". *Nature Communications* 2021 12:1 12: pp. 1–12.
- Burnham, K. L., Davenport, E. E., Radhakrishnan, J., et al. (2017). "Shared and Distinct Aspects of the Sepsis Transcriptomic Response to Fecal Peritonitis and Pneumonia". *American journal of respiratory and critical care medicine* 196: pp. 328–339.
- Calis, J. J. and Rosenberg, B. R. (2014). "Characterizing immune repertoires by high throughput sequencing: strategies and applications". *Trends in immunology* 35: p. 581.
- Cameron, M. J., Ran, L., Xu, L., et al. (2007). "Interferon-Mediated Immunopathological Events Are Associated with Atypical Innate and Adaptive Immune Responses in Patients with Severe Acute Respiratory Syndrome". *Journal of Virology* 81: pp. 8692–8706.

- Cano-Gamez, E., Burnham, K. L., Goh, C., et al. (2022). "An immune dysfunction score for stratification of patients with acute infection based on whole blood gene expression". *medRxiv*: p. 2022.03.17.22272427.
- Cano-Gamez, E., Soskic, B., Roumeliotis, T. I., et al. (2020). "Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines". *Nature Communications* 2020 11:1 11: pp. 1–15.
- Caricchio, R., Abbate, A., Gordeev, I., et al. (2021). "Effect of Canakinumab vs Placebo on Survival Without Invasive Mechanical Ventilation in Patients Hospitalized With Severe COVID-19: A Randomized Clinical Trial". *JAMA* 326: pp. 230–239.
- Carsana, L., Sonzogni, A., Nasr, A., et al. (2020). "Pulmonary post-mortem findings in a series of COVID-19 cases from northern Italy: a two-centre descriptive study". *The Lancet. Infectious diseases* 20: pp. 1135–1140.
- Casanova, J. L. and Abel, L. (2021). "Lethal Infectious Diseases as Inborn Errors of Immunity: Toward a Synthesis of the Germ and Genetic Theories". *Annual review of pathology* 16: pp. 23–50.
- Cavaillon, J. M. and Annane, D. (2006). "Compartmentalization of the inflammatory response in sepsis and SIRS". *Journal of endotoxin research* 12: pp. 151–170.
- Cavalli, G., De Luca, G., Campochiaro, C., et al. (2020). "Interleukin-1 blockade with high-dose anakinra in patients with COVID-19, acute respiratory distress syndrome, and hyperinflammation: a retrospective cohort study". *The Lancet. Rheumatology* 2: e325–e331.
- Chalise, P. and Fridley, B. L. (2017). "Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm". *PLOS ONE* 12: e0176278.
- Chang, K., Svabek, C., Vazquez-Guillamet, C., et al. (2014). "Targeting the programmed cell death 1: programmed cell death ligand 1 pathway reverses T cell exhaustion in patients with sepsis". *Critical care (London, England)* 18.
- Chang, K. C., Unsinger, J., Davis, C. G., et al. (2007). "Multiple triggers of cell death in sepsis: death receptor and mitochondrial-mediated apoptosis". *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 21: pp. 708–719.
- Chaussabel, D. and Baldwin, N. (2014). "Democratizing systems immunology with modular transcriptional repertoire analyses". *Nature Reviews Immunology* 2014 14:4 14: pp. 271–280.
- Chaussabel, D., Quinn, C., Shen, J., et al. (2008). "A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus". *Immunity* 29: pp. 150–164.
- Chen, L., Deng, H., Cui, H., et al. (2018). "Inflammatory responses and inflammation-associated diseases in organs". *Oncotarget* 9: p. 7204.
- Chen, L. Y. and Quach, T. T. (2021). "COVID-19 cytokine storm syndrome: a threshold concept". *The Lancet Microbe* 2: e49–e50.
- Chen, Z. and John Wherry, E. (2020). "T cell responses in patients with COVID-19". *Nature Reviews Immunology* 2020 20:9 20: pp. 529–536.

- Cheng, S. C., Scicluna, B. P., Arts, R. J., et al. (2016). "Broad defects in the energy metabolism of leukocytes underlie immunoparalysis in sepsis". *Nature immunology* 17: pp. 406–413.
- Chertow, D., Stein, S., Ramelli, S., et al. (2021). "SARS-CoV-2 infection and persistence throughout the human body and brain".
- Chiang, M. K. and Melton, D. A. (2003). "Single-Cell Transcript Analysis of Pancreas Development". *Developmental Cell* 4: pp. 383–393.
- Chu, T., Wang, Z., Pe'er, D., et al. (2022). "Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology". *Nature Cancer* 2022 3:4 3: pp. 505–517.
- Chua, R. L., Lukassen, S., Trump, S., et al. (2020). "COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis". *Nature biotechnology* 38: pp. 970–979.
- Ciancanelli, M. J., Huang, S. X., Luthra, P., et al. (2015). "Infectious disease. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency". *Science (New York, N.Y.)* 348: pp. 448–453.
- Cilloniz, C., Peroni, H. J., Gabarrús, A., et al. (2021). "Lymphopenia Is Associated With Poor Outcomes of Patients With Community-Acquired Pneumonia and Sepsis". *Open Forum Infectious Diseases* 8.
- Clark, S. J., Argelaguet, R., Kapourani, C. A., et al. (2018). "scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells". *Nature Communications* 2018 9:1 9: pp. 1–9.
- Clauss, M., Weich, H., Breier, G., et al. (1996). "The vascular endothelial growth factor receptor Flt-1 mediates biological activities. Implications for a functional role of placenta growth factor in monocyte activation and chemotaxis". *The Journal of biological chemistry* 271: pp. 17629–17634.
- Coifman, R. R. and Lafon, S. (2006). "Diffusion maps". *Appl. Comput. Harmon. Anal* 21: pp. 5–30.
- COMBAT consortium (2022). "A blood atlas of COVID-19 defines hallmarks of disease severity and specificity". *Cell* 185: 916–938.e58.
- Condamine, T., Dominguez, G. A., Youn, J. I., et al. (2016). "Lectin-type oxidized LDL receptor-1 distinguishes population of human polymorphonuclear myeloid-derived suppressor cells in cancer patients". *Science immunology* 1.
- Conesa, A., Madrigal, P., Tarazona, S., et al. (2016). "A survey of best practices for RNA-seq data analysis". *Genome Biology* 2016 17:1 17: pp. 1–19.
- Conway Morris, A., Datta, D., Shankar-Hari, M., et al. (2018). "Cell-surface signatures of immune dysfunction risk-stratify critically ill patients: INFECT study". *Intensive Care Medicine* 44: p. 627.
- Conway Morris, A., Kefala, K., Wilkinson, T. S., et al. (2009). "C5a mediates peripheral blood neutrophil dysfunction in critically ill patients". *American journal of respiratory and critical care medicine* 180: pp. 19–28.
- COVID-19 Forecasting Team (2022). "Variation in the COVID-19 infection–fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis". *The Lancet* 399: pp. 1469–1488.

- Cozza, V., Campbell, H., Chang, H. H., et al. (2021). "Global Seasonal Influenza Mortality Estimates: A Comparison of 3 Different Approaches". *American journal of epidemiology* 190: p. 718.
- Crotty, S. (2015). "A brief history of T cell help to B cells". *Nature Reviews Immunology* 2015 15:3 15: pp. 185–189.
- Crowell, H. L., Soneson, C., Germain, P. L., et al. (2020). "muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data". *Nature Communications* 2020 11:1 11: pp. 1–12.
- Cruikshank, W. W., Kornfeld, H., and Center, D. M. (2000). "Interleukin-16". *Journal of leukocyte biology* 67: pp. 757–766.
- Cuthbertson, B. H., Elders, A., Hall, S., et al. (2013). "Mortality and quality of life in the five years after severe sepsis". *Critical Care* 17: pp. 1–8.
- Daix, T., Guerin, E., Tavernier, E., et al. (2018). "Multicentric Standardized Flow Cytometry Routine Assessment of Patients With Sepsis to Predict Clinical Worsening". *Chest* 154: pp. 617–627.
- Danek, P., Kardosova, M., Janeckova, L., et al. (2020). " $\beta$ -Catenin-TCF/LEF signaling promotes steady-state and emergency granulopoiesis via G-CSF receptor upregulation". *Blood* 136: pp. 2574–2587.
- Dann, E., Henderson, N. C., Teichmann, S. A., et al. (2021). "Differential abundance testing on single-cell data using k-nearest neighbor graphs". *Nature Biotechnology* 2021 40:2 40: pp. 245–253.
- Davenport, E. E., Burnham, K. L., Radhakrishnan, J., et al. (2016). "Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study". *The Lancet. Respiratory medicine* 4: pp. 259–271.
- Delaloye, J. and Calandra, T. (2014). "Invasive candidiasis as a cause of sepsis in the critically ill patient". *Virulence* 5: p. 161.
- Delano, M. J., Scumpia, P. O., Weinstein, J. S., et al. (2007). "MyD88-dependent expansion of an immature GR-1(+)CD11b(+) population induces T cell suppression and Th2 polarization in sepsis". *The Journal of experimental medicine* 204: pp. 1463–1474.
- Demerle, K. M., Angus, D. C., Baillie, J. K., et al. (2021). "Sepsis Subclasses: A Framework for Development and Interpretation". *Critical care medicine* 49: pp. 748–759.
- Demers, M., Krause, D. S., Schatzberg, D., et al. (2012). "Cancers predispose neutrophils to release extracellular DNA traps that contribute to cancer-associated thrombosis". *Proceedings of the National Academy of Sciences of the United States of America* 109: pp. 13076–13081.
- Denning, N. L., Aziz, M., Gurien, S. D., et al. (2019). "Damps and nets in sepsis". *Frontiers in Immunology* 10: p. 2536.
- Derde, L. P. G., Gordon, A. C., Mouncey, P. R., et al. (2021). "Effectiveness of Tocilizumab, Sarilumab, and Anakinra for critically ill patients with COVID-19 The REMAP-CAP COVID-19 Immune Modulation Therapy Domain Randomized Clinical Trial". *medRxiv*: p. 2021.06.18.21259133.
- Deutschman, C. S. and Tracey, K. J. (2014). "Sepsis: Current Dogma and New Perspectives". *Immunity* 40: pp. 463–475.

- Devi, S., Wang, Y., Chew, W. K., et al. (2013). "Neutrophil mobilization via plerixafor-mediated CXCR4 inhibition arises from lung demargination and blockade of neutrophil homing to the bone marrow". *The Journal of Experimental Medicine* 210: p. 2321.
- Dietlin, T. A., Hofman, F. M., Lund, B. T., et al. (2007). "Mycobacteria-induced Gr-1+ subsets from distinct myeloid lineages have opposite effects on T cell expansion". *Journal of leukocyte biology* 81: pp. 1205–1212.
- Diggins, K. E., Greenplate, A. R., Leelatian, N., et al. (2017). "Characterizing cell subsets using marker enrichment modeling". *Nature Methods* 2017 14:3 14: pp. 275–278.
- Dobin, A., Davis, C. A., Schlesinger, F., et al. (2013). "STAR: ultrafast universal RNA-seq aligner". *Bioinformatics* 29: pp. 15–21.
- Dong, M., Thennavan, A., Urrutia, E., et al. (2021). "SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references". *Briefings in Bioinformatics* 22: pp. 416–427.
- Donohue, D. M., Reiff, R. H., Hanson, M. L., et al. (1958). "Quantitative measurement of the erythrocytic and granulocytic cells of the marrow and blood". *The Journal of clinical investigation* 37: pp. 1571–1576.
- Drewry, A., Samra, N., Skrupky, L., et al. (2014). "Persistent Lymphopenia after Diagnosis of Sepsis Predicts Mortality". *Shock (Augusta, Ga.)* 42: p. 383.
- Engelmann, B. and Massberg, S. (2013). "Thrombosis as an intravascular effector of innate immunity". *Nature reviews. Immunology* 13: pp. 34–45.
- Fabregat, A., Jupe, S., Matthews, L., et al. (2018). "The Reactome Pathway Knowledgebase". *Nucleic Acids Research* 46: pp. D649–D655.
- Fang, H., Knezevic, B., Burnham, K. L., et al. (2016). "XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits". *Genome Medicine* 8: pp. 1–20.
- Fisher, C. J., Dhainaut, J.-F. A., Opal, S. M., et al. (1994). "Recombinant Human Interleukin 1 Receptor Antagonist in the Treatment of Patients With Sepsis Syndrome: Results From a Randomized, Double-blind, Placebo-Controlled Trial". *JAMA* 271: pp. 1836–1843.
- Fleischmann-Struzek, C., Mellhammar, L., Rose, N., et al. (2020). "Incidence and mortality of hospital- and ICU-treated sepsis: results from an updated and expanded systematic review and meta-analysis". *Intensive Care Medicine* 46: pp. 1552–1562.
- Fleischmann-Struzek, C., Rose, N., Freytag, A., et al. (2021). "Epidemiology and Costs of Postsepsis Morbidity, Nursing Care Dependency, and Mortality in Germany, 2013 to 2017". *JAMA Network Open* 4: e2134290–e2134290.
- Flerlage, T., Boyd, D. F., Meliopoulos, V., et al. (2021). "Influenza virus and SARS-CoV-2: pathogenesis and host responses in the respiratory tract". *Nature Reviews Microbiology* 2021 19:7 19: pp. 425–441.
- Furze, R. C. and Rankin, S. M. (2008). "Neutrophil mobilization and clearance in the bone marrow". *Immunology* 125: p. 281.
- Gabrilovich, D. I., Bronte, V., Chen, S. H., et al. (2007). "The terminology issue for myeloid-derived suppressor cells". *Cancer research* 67: p. 425.
- Gaebler, C., Wang, Z., Lorenzi, J. C., et al. (2021). "Evolution of antibody immunity to SARS-CoV-2". *Nature* 591: pp. 639–644.



- Gao, C., Zhang, M., and Chen, L. (2020). "The Comparison of Two Single-cell Sequencing Platforms: BD Rhapsody and 10x Genomics Chromium". *Current genomics* 21: pp. 602–609.
- Gayoso, A., Lopez, R., Xing, G., et al. (2022). "A Python library for probabilistic analysis of single-cell omics data". *Nature Biotechnology* 2022 40:2 40: pp. 163–166.
- Gayoso, A., Steier, Z., Lopez, R., et al. (2021). "Joint probabilistic modeling of single-cell multi-omic data with totalVI". *Nature Methods* 2021 18:3 18: pp. 272–282.
- Giladi, A., Paul, F., Herzog, Y., et al. (2018). "Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis". *Nature Cell Biology* 2018 20:7 20: pp. 836–846.
- Godfrey, D. I., Uldrich, A. P., McCluskey, J., et al. (2015). "The burgeoning family of unconventional T cells". *Nature Immunology* 2015 16:11 16: pp. 1114–1123.
- Goh, C., Golubchik, T., Ansari, M. A., et al. (2019). "Targeted metagenomic sequencing enhances the identification of pathogens associated with acute infection". *bioRxiv*: p. 716902.
- Goh, C. and Knight, J. C. (2017). "Enhanced understanding of the host-pathogen interaction in sepsis: new opportunities for omic approaches". *The Lancet. Respiratory medicine* 5: pp. 212–223.
- Graier, J. J., Kalbitz, M., Zetoune, F. S., et al. (2014). "Persistent Neutrophil Dysfunction and Suppression of Acute Lung Injury in Mice following Cecal Ligation and Puncture Sepsis". *Journal of Innate Immunity* 6: pp. 695–705.
- Grimes, J. M. and Grimes, K. V. (2020). "p38 MAPK inhibition: A promising therapeutic approach for COVID-19". *Journal of Molecular and Cellular Cardiology* 144: p. 63.
- Group, T. S. C.-. G. (2020). "Genomewide Association Study of Severe Covid-19 with Respiratory Failure". *New England Journal of Medicine* 383: pp. 1522–1534.
- Guan, W.-j., Ni, Z.-y., Hu, Y., et al. (2020). "Clinical Characteristics of Coronavirus Disease 2019 in China". *The New England journal of medicine* 382: pp. 1708–1720.
- Guigo, R. and Hoon, M. de (2018). "Recent advances in functional genome analysis". *F1000Research* 7.
- Hall, C. J., Flores, M. V., Oehlers, S. H., et al. (2012). "Infection-Responsive Expansion of the Hematopoietic Stem and Progenitor Cell Compartment in Zebrafish Is Dependent upon Inducible Nitric Oxide". *Cell Stem Cell* 10: pp. 198–209.
- Hao, Y., Hao, S., Andersen-Nissen, E., et al. (2021). "Integrated analysis of multimodal single-cell data". *Cell* 184: 3573–3587.e29.
- Hao, Y., Stuart, T., Kowalski, M., et al. (2022). "Dictionary learning for integrative, multimodal, and scalable single-cell analysis". *bioRxiv*: p. 2022.02.24.481684.
- Harris, H. (2000). *The Birth of the Cell*. New Haven: Yale University Press: p. 224.
- Hay, S. B., Ferchen, K., Chetal, K., et al. (2018). "The Human Cell Atlas bone marrow single-cell interactive web portal". *Experimental Hematology* 68: pp. 51–61.

- Hernandez, N., Melki, I., Jing, H., et al. (2018). "Life-threatening influenza pneumonitis in a child with inherited IRF9 deficiency". *The Journal of experimental medicine* 215: pp. 2567–2585.
- Hernandez-Beeftink, T., Guillen-Guio, B., Lorenzo-Salazar, J. M., et al. (2022). "A Genome-Wide Association Study of Survival in Patients with Sepsis". *medRxiv* 19: p. 2022.05.06.22274756.
- Hikmet, F., Méar, L., Edvinsson, Å., et al. (2020). "The protein expression profile of ACE2 in human tissues". *Molecular Systems Biology* 16: e9610.
- Hirai, H., Zhang, P., Dayaram, T., et al. (2006). "C/EBPbeta is required for 'emergency' granulopoiesis". *Nature immunology* 7: pp. 732–739.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., et al. (2020). "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor". *Cell* 181: 271–280.e8.
- Höllt, T., Pezzotti, N., Unen, V. van, et al. (2016). "Cytosplore: Interactive Immune Cell Phenotyping for Large Single-Cell Datasets". *Computer Graphics Forum* 35: pp. 171–180.
- Honda, K., Takaoka, A., and Taniguchi, T. (2006). "Type I Inteferon Gene Induction by the Interferon Regulatory Factor Family of Transcription Factors". *Immunity* 25: pp. 349–360.
- Hoogerwerf, J. J., De Vos, A. F., Van't Veer, C., et al. (2010). "Priming of alveolar macrophages upon instillation of lipopolysaccharide in the human lung". *American journal of respiratory cell and molecular biology* 42: pp. 349–356.
- Hope, C. M., Huynh, D., Wong, Y. Y., et al. (2021). "Optimization of blood handling and peripheral blood mononuclear cell cryopreservation of low cell number samples". *International Journal of Molecular Sciences* 22: p. 9129.
- Hore, V., Viñuela, A., Buil, A., et al. (2016). "Tensor decomposition for multi-tissue gene expression experiments". *Nature genetics* 48: p. 1094.
- Horowitz, J. E., Kosmicki, J. A., Damask, A., et al. (2022). "Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease". *Nature Genetics* 2022 54:4 54: pp. 382–392.
- Hotchkiss, R. S., Moldawer, L. L., Opal, S. M., et al. (2016). "Sepsis and septic shock". *Nature Reviews Disease Primers* 2016 2:1 2: pp. 1–21.
- Hotchkiss, R. S., Swanson, P. E., Freeman, B. D., et al. (1999). "Apoptotic cell death in patients with sepsis, shock, and multiple organ dysfunction". *Critical care medicine* 27: pp. 1230–1251.
- Hou, Y. J., Okuda, K., Edwards, C. E., et al. (2020). "SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract". *Cell* 182: 429–446.e14.
- Huang, C., Wang, Y., Li, X., et al. (2020). "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China". *The Lancet* 395: pp. 497–506.
- Huang, Y., McCarthy, D. J., and Stegle, O. (2019). "Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference". *Genome Biology* 20: pp. 1–12.
- Huertas, A., Montani, D., Savale, L., et al. (2020). "Endothelial cell dysfunction: a major player in SARS-CoV-2 infection (COVID-19)?" *European Respiratory Journal* 56.

- Huet, T., Beaussier, H., Voisin, O., et al. (2020). "Anakinra for severe forms of COVID-19: a cohort study". *The Lancet. Rheumatology* 2: e393–e400.
- Hyun, K. K., De La Luz Sierra, M., Williams, C. K., et al. (2006). "G-CSF down-regulation of CXCR4 expression identified as a mechanism for mobilization of myeloid cells". *Blood* 108: pp. 812–820.
- Iacobucci, G. (2021). "Covid and flu: what do the numbers tell us about morbidity and deaths?" *BMJ* 375: n2514.
- Ince, C., Mayeux, P. R., Nguyen, T., et al. (2016). "THE ENDOTHELIUM IN SEPSIS". *Shock (Augusta, Ga.)* 45: p. 259.
- Janky, R., Verfaillie, A., Imrichová, H., et al. (2014). "iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections". *PLOS Computational Biology* 10: e1003731.
- Jiang, R., Sun, T., Song, D., et al. (2022). "Statistics or biology: the zero-inflation controversy about scRNA-seq data". *Genome Biology* 2022 23:1 23: pp. 1–24.
- Jovic, D., Liang, X., Zeng, H., et al. (2022). "Single-cell RNA sequencing technologies and applications: A brief overview". *Clinical and Translational Medicine* 12: e694.
- Kamme, F., Salunga, R., Yu, J., et al. (2003). "Single-Cell Microarray Analysis in Hippocampus CA1: Demonstration and Validation of Cellular Heterogeneity". *Journal of Neuroscience* 23: pp. 3607–3615.
- Karakike, E., Giamarellos-Bourboulis, E. J., Kyprianou, M., et al. (2021). "Coronavirus Disease 2019 as Cause of Viral Sepsis: A Systematic Review and Meta-Analysis". *Critical care medicine* 49: pp. 2042–2057.
- Kedl, R. M. and Tamburini, B. A. (2015). "Antigen archiving by lymph node stroma: A novel function for the lymphatic endothelium". *European journal of immunology* 45: pp. 2721–2729.
- Kharchenko, P. V., Silberstein, L., and Scadden, D. T. (2014). "Bayesian approach to single-cell differential expression analysis". *Nature methods* 11: p. 740.
- Kim, W. Y. and Hong, S. B. (2016). "Sepsis and Acute Respiratory Distress Syndrome: Recent Update". *Tuberculosis and Respiratory Diseases* 79: p. 53.
- Kooistra, E. J., Waalders, N. J., Grondman, I., et al. (2020). "Anakinra treatment in critically ill COVID-19 patients: a prospective cohort study". *Critical care (London, England)* 24.
- Korsunsky, I., Millard, N., Fan, J., et al. (2019). "Fast, sensitive and accurate integration of single-cell data with Harmony". *Nature Methods* 2019 16:12 16: pp. 1289–1296.
- Köstlin, N., Vogelmann, M., Spring, B., et al. (2017). "Granulocytic myeloid-derived suppressor cells from human cord blood modulate T-helper cell response towards an anti-inflammatory phenotype". *Immunology* 152: pp. 89–101.
- Kotfis, K., Wittebole, X., Jaschinski, U., et al. (2019). "A worldwide perspective of sepsis epidemiology and survival according to age: Observational data from the ICON audit". *Journal of critical care* 51: pp. 122–132.
- Kotliar, D., Veres, A., Nagy, M. A., et al. (2019). "Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq". *eLife* 8.



- Kousathanas, A., Pairo-Castineira, E., Rawlik, K., et al. (2022). "Whole-genome sequencing reveals host factors underlying critical COVID-19". *Nature* 2022: pp. 1–16.
- Kox, M., Waalders, N. J., Kooistra, E. J., et al. (2020). "Cytokine Levels in Critically Ill Patients With COVID-19 and Other Conditions". *JAMA* 324: pp. 1565–1567.
- Krammer, F., Smith, G. J., Fouchier, R. A., et al. (2018). "Influenza". *Nature Reviews Disease Primers* 2018 4:1 4: pp. 1–21.
- Kurt-Jones, E. A., Beller, D. I., Mizel, S. B., et al. (1985). "Identification of a membrane-associated interleukin 1 in macrophages". *Proceedings of the National Academy of Sciences of the United States of America* 82: pp. 1204–1208.
- Kusnadi, A., Ramírez-Suástegui, C., Fajardo, V., et al. (2021). "Severely ill COVID-19 patients display impaired exhaustion features in SARS-CoV-2-reactive CD8+ T cells". *Science Immunology* 6: p. 4782.
- Kwok, A. J., Mentzer, A., and Knight, J. C. (2021). "Host genetics and infectious disease: new tools, insights and translational opportunities". *Nature reviews. Genetics* 22: pp. 137–153.
- Landelle, C., Lepape, A., Voirin, N., et al. (2010). "Low monocyte human leukocyte antigen-DR is independently associated with nosocomial infections after septic shock". *Intensive care medicine* 36: pp. 1859–1866.
- Langfelder, P. and Horvath, S. (2008). "WGCNA: An R package for weighted correlation network analysis". *BMC Bioinformatics* 9: pp. 1–13.
- Law, C. W., Chen, Y., Shi, W., et al. (2014). "Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts". *Genome Biology* 15: pp. 1–17.
- Lee, A. J. and Ashkar, A. A. (2018). "The dual nature of type I and type II interferons". *Frontiers in Immunology* 9: p. 2061.
- Lee, J. S., Park, S., Jeong, H. W., et al. (2020). "Immunophenotyping of covid-19 and influenza highlights the role of type i interferons in development of severe covid-19". *Science Immunology* 5: p. 1554.
- Leek, J. T., Johnson, W. E., Parker, H. S., et al. (2012). "The sva package for removing batch effects and other unwanted variation in high-throughput experiments". *Bioinformatics* 28: p. 882.
- Leligdowicz, A. and Matthay, M. A. (2019). "Heterogeneity in sepsis: New biological evidence with clinical applications". *Critical Care* 23: pp. 1–8.
- Levy, M. M., Fink, M. P., Marshall, J. C., et al. (2003). "2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference". *Critical care medicine* 31: pp. 1250–1256.
- Li, D. and Wu, M. (2021). "Pattern recognition receptors in health and diseases". *Signal transduction and targeted therapy* 6.
- Liao, M., Liu, Y., Yuan, J., et al. (2020). "Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19". *Nature Medicine* 2020 26:6 26: pp. 842–844.
- Liao, Y., Smyth, G. K., and Shi, W. (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". *Bioinformatics* 30: pp. 923–930. arXiv: 1305.3347.
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., et al. (2015). "The Molecular Signatures Database (MSigDB) hallmark gene set collection". *Cell systems* 1: p. 417.

- Lim, H. K., Huang, S. X., Chen, J., et al. (2019). "Severe influenza pneumonitis in children with inherited TLR3 deficiency". *The Journal of experimental medicine* 216: pp. 2038–2056.
- Lindeboom, R. G., Regev, A., and Teichmann, S. A. (2021). "Towards a Human Cell Atlas: Taking Notes from the Past". *Trends in Genetics* 37: pp. 625–630.
- Ling, V., Luxenberg, D., Wang, J., et al. (1997). "Structural identification of the hematopoietic progenitor antigen ER-MP12 as the vascular endothelial adhesion molecule PECAM-1 (CD31)". *European journal of immunology* 27: pp. 509–514.
- Lippi, G. and Plebani, M. (2020). "Laboratory abnormalities in patients with COVID-2019 infection". *Clinical Chemistry and Laboratory Medicine* 58: pp. 1131–1134.
- Loo, J., Spittle, D. A., and Newnham, M. (2021). "COVID-19, immunothrombosis and venous thromboembolism: biological mechanisms". *Thorax* 76: pp. 412–420.
- Lopez, R., Regier, J., Cole, M. B., et al. (2018). "Deep generative modeling for single-cell transcriptomics". *Nature Methods* 15:12 15: pp. 1053–1058.
- Love, M. I., Huber, W., and Anders, S. (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biology* 15: pp. 1–21.
- Lozano, R., Naghavi, M., Foreman, K., et al. (2012). "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010". *Lancet (London, England)* 380: pp. 2095–2128.
- Lu, L. L., Suscovich, T. J., Fortune, S. M., et al. (2017). "Beyond binding: antibody effector functions in infectious diseases". *Nature Reviews Immunology* 17:18 18: pp. 46–61.
- Lucas, C., Wong, P., Klein, J., et al. (2020). "Longitudinal analyses reveal immunological misfiring in severe COVID-19". *Nature* 2020 584:7821 584: pp. 463–469.
- Lun, A. T., Riesenfeld, S., Andrews, T., et al. (2019). "EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data". *Genome Biology* 20: pp. 1–9.
- Luo, B., Yu, Z., and Li, Y. (2017). "Thyroid hormone disorders and sepsis". *Bio-medical materials and engineering* 28: S237–S241.
- Ma, S., Zhang, B., LaFave, L. M., et al. (2020). "Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin". *Cell* 183: 1103–1116.e20.
- Macosko, E. Z., Basu, A., Satija, R., et al. (2015). "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets". *Cell* 161: pp. 1202–1214.
- Mahase, E. (2020). "Covid-19: death rate is 0.66estimates". *BMJ* 369.
- Maji, A., Misra, R., Mondal, A. K., et al. (2015). "Expression profiling of lymph nodes in tuberculosis patients reveal inflammatory milieu at site of infection". *Scientific Reports* 2015 5:1 5: pp. 1–10.
- Malas, M. B., Naazie, I. N., Elsayed, N., et al. (2020). "Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: A systematic review and meta-analysis". *EClinicalMedicine* 29-30: p. 100639.

- Manz, M. G. and Boettcher, S. (2014). "Emergency granulopoiesis". *Nature Reviews Immunology* 2014 14:5 14: pp. 302–314.
- Marcenaro, E., Notarangelo, L. D., Orange, J. S., et al. (2017). "Editorial: NK Cell Subsets in Health and Disease: New Developments". *Frontiers in Immunology* 8.
- Mariette, X., Hermine, O., Resche-Rigon, M., et al. (2021). "Effect of anakinra versus usual care in adults in hospital with COVID-19 and mild-to-moderate pneumonia (CORIMUNO-ANA-1): a randomised controlled trial". *The Lancet. Respiratory medicine* 9: pp. 295–304.
- Marini, O., Spina, C., Mimiola, E., et al. (2016). "Identification of granulocytic myeloid-derived suppressor cells (G-MDSCs) in the peripheral blood of Hodgkin and non-Hodgkin lymphoma patients". *Oncotarget* 7: pp. 27676–27688.
- Marques-Coelho, D., Iohan, L. d. C. C., Melo de Farias, A. R., et al. (2021). "Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains". *npj Aging and Mechanisms of Disease* 2021 7:1 7: pp. 1–15.
- Marshall, J. C. (2008). "Sepsis: rethinking the approach to clinical research". *Journal of Leukocyte Biology* 83: pp. 471–482.
- Marshall, J. C. (2014). "Why have clinical trials in sepsis failed?" *Trends in Molecular Medicine* 20: pp. 195–203.
- Mauer, A. M., Athens, J. W., Ashenbrucker, H., et al. (1960). "LEUKOKINETIC STUDIES. II. A METHOD FOR LABELING GRANULOCYTES IN VITRO WITH RADIOACTIVE DIISOPROPYLFLUOROPHOSPHATE (DFP32)". *Journal of Clinical Investigation* 39: p. 1481.
- Maurano, M. T., Humbert, R., Rynes, E., et al. (2012). "Systematic Localization of Common Disease-Associated Variation in Regulatory DNA". *Science (New York, N.Y.)* 337: p. 1190.
- Mayassi, T., Barreiro, L. B., Rossjohn, J., et al. (2021). "A multilayered immune system through the lens of unconventional T cells". *Nature* 595: pp. 501–510.
- Mayr, F. B., Yende, S., and Angus, D. C. (2014). "Epidemiology of severe sepsis". *Virulence* 5: p. 4.
- McClure, C., Ali, E., Youssef, D., et al. (2016). "NFI-A disrupts myeloid cell differentiation and maturation in septic mice". *Journal of leukocyte biology* 99: pp. 201–211.
- McClure, C., Brudecki, L., Ferguson, D. A., et al. (2014). "MicroRNA 21 (miR-21) and miR-181b couple with NFI-A to generate myeloid-derived suppressor cells and promote immunosuppression in late sepsis". *Infection and immunity* 82: pp. 3816–3825.
- McClure, C., McPeak, M. B., Youssef, D., et al. (2017). "Stat3 and C/EBP $\beta$  synergize to induce miR-21 and miR-181b expression during sepsis". *Immunology and cell biology* 95: pp. 42–55.
- McKinnon, K. M. (2018). "Flow Cytometry: An Overview". *Current protocols in immunology* 120: p. 5.1.1.
- McNab, F., Mayer-Barber, K., Sher, A., et al. (2015). "Type I interferons in infectious disease". *Nature Reviews Immunology* 2015 15:2 15: pp. 87–103.

- McPeak, M. B., Youssef, D., Williams, D. A., et al. (2017). "Frontline Science: Myeloid cell-specific deletion of *Cebpb* decreases sepsis-induced immunosuppression in mice". *Journal of Leukocyte Biology* 102: pp. 191–200.
- Metzemaekers, M., Gouwy, M., and Proost, P. (2020). "Neutrophil chemoattractant receptors in health and disease: double-edged swords". *Cellular Molecular Immunology* 2020 17:5 17: pp. 433–450.
- Meyerowitz-Katz, G. and Merone, L. (2020). "A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates". *International Journal of Infectious Diseases* 101: pp. 138–148.
- Monneret, G., Lepape, A., Voirin, N., et al. (2006). "Persisting low monocyte human leukocyte antigen-DR expression predicts mortality in septic shock". *Intensive care medicine* 32: pp. 1175–1183.
- Moroz, L. L., Edwards, J. R., Puthanveetil, S. V., et al. (2006). "Neuronal Transcriptome of *Aplysia*: Neuronal Compartments and Circuitry". *Cell* 127: pp. 1453–1467.
- Morris, A. C., Brittan, M., Wilkinson, T. S., et al. (2011). "C5a-mediated neutrophil dysfunction is RhoA-dependent and predicts infection in critically ill patients". *Blood* 117: pp. 5178–5188.
- Mulè, M. P., Martins, A. J., and Tsang, J. S. (2022). "Normalizing and denoising protein expression data from droplet-based single cell profiling". *Nature Communications* 2022 13:1 13: pp. 1–12.
- Murphy, K. (2017). *Janeway's immunobiology*. 9th editio. New York: Garland Science: pp. 1–277.
- Muus, C., Luecken, M. D., Eraslan, G., et al. (2021). "Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics". *Nature Medicine* 2021 27:3 27: pp. 546–559.
- Nagai, Y., Garrett, K. P., Ohta, S., et al. (2006). "Toll-like receptors on hematopoietic progenitor cells stimulate innate immune system replenishment". *Immunity* 24: pp. 801–812.
- Nalbandian, A., Sehgal, K., Gupta, A., et al. (2021). "Post-acute COVID-19 syndrome". *Nature Medicine* 2021 27:4 27: pp. 601–615.
- Natarajan, A., Zlitni, S., Brooks, E. F., et al. (2022). "Gastrointestinal symptoms and fecal shedding of SARS-CoV-2 RNA suggest prolonged gastrointestinal infection". *Med* 0.
- Nathan, A., Asgari, S., Ishigaki, K., et al. (2022). "Single-cell eQTL models reveal dynamic T cell state dependence of disease loci". *Nature* 2022 606:7912 606: pp. 120–128.
- Newman, A. M., Liu, C. L., Green, M. R., et al. (2015). "Robust enumeration of cell subsets from tissue expression profiles". *Nature Methods* 2015 12:5 12: pp. 453–457.
- Newman, A. M., Steen, C. B., Liu, C. L., et al. (2019). "Determining cell type abundance and expression from bulk tissues with digital cytometry". *Nature Biotechnology* 2019 37:7 37: pp. 773–782.
- Niemi, M. E., Karjalainen, J., Liao, R. G., et al. (2021). "Mapping the human genetic architecture of COVID-19". *Nature* 2021 600:7889 600: pp. 472–477.
- Nordenfelt, P. and Tapper, H. (2011). "Phagosome dynamics during phagocytosis by neutrophils". *Journal of leukocyte biology* 90: pp. 271–284.

- Nowicka, M., Krieg, C., Crowell, H. L., et al. (2019). "CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets". *F1000Research* 2019 6:748 6: p. 748.
- Obi, A. T., Barnes, G. D., Napolitano, L. M., et al. (2021). "Venous thrombosis epidemiology, pathophysiology, and anticoagulant therapies and trials in severe acute respiratory syndrome coronavirus 2 infection". *Journal of vascular surgery. Venous and lymphatic disorders* 9: pp. 23–35.
- Obi, A. T., Pannucci, C. J., Nackashi, A., et al. (2015). "Validation of the Caprini Venous Thromboembolism Risk Assessment Model in Critically Ill Surgical Patients". *JAMA surgery* 150: pp. 941–948.
- O'Connell, M. J. and Lock, E. F. (2016). "R.JIVE for exploration of multi-source molecular data". *Bioinformatics* 32: pp. 2877–2879.
- Ogishi, M., Yang, R., Gruber, C., et al. (2021). "Multibatch Cytometry Data Integration for Optimal Immunophenotyping". *Journal of immunology (Baltimore, Md. : 1950)* 206: pp. 206–213.
- Opal, S. M. and Poll, T. van der (2015). "Endothelial barrier dysfunction in septic shock". *Journal of Internal Medicine* 277: pp. 277–293.
- Opal, S. M., Fisher, C. J., Dhainaut, J. F. A., et al. (1997). "Confirmatory interleukin-1 receptor antagonist trial in severe sepsis: a phase III, randomized, double-blind, placebo-controlled, multicenter trial. The Interleukin-1 Receptor Antagonist Sepsis Investigator Group". *Critical care medicine* 25: pp. 1115–1124.
- Osuchowski, M. F., Winkler, M. S., Skirecki, T., et al. (2021). "The COVID-19 puzzle: deciphering pathophysiology and phenotypes of a new disease entity". *The Lancet Respiratory Medicine* 9: pp. 622–642.
- Pairo-Castineira, E., Clohisey, S., Klaric, L., et al. (2020). "Genetic mechanisms of critical illness in COVID-19". *Nature* 2020 591:7848 591: pp. 92–98.
- Palanichamy, A., Barnard, J., Zheng, B., et al. (2009). "Novel human transitional B cell populations revealed by B cell depletion therapy". *Journal of immunology (Baltimore, Md. : 1950)* 182: p. 5982.
- Papayannopoulos, V. (2017). "Neutrophil extracellular traps in immunity and disease". *Nature Reviews Immunology* 2017 18:2 18: pp. 134–147.
- Patra, T., Meyer, K., Geerling, L., et al. (2020). "SARS-CoV-2 spike protein promotes IL-6 trans-signaling by activation of angiotensin II receptor signaling in epithelial cells". *PLOS Pathogens* 16: e1009128.
- Pauken, K. E., Sammons, M. A., Odorizzi, P. M., et al. (2016). "Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade". *Science (New York, N.Y.)* 354: pp. 1160–1165.
- Pavlov, V. A., Chavan, S. S., and Tracey, K. J. (2018). "Molecular and Functional Neuroscience in Immunity". *Annual review of immunology* 36: pp. 783–812.
- Pavlov, V. A. and Tracey, K. J. (2012). "The vagus nerve and the inflammatory reflex—linking immunity and metabolism". *Nature reviews. Endocrinology* 8: p. 743.
- Pena, O. M., Hancock, D. G., Lyle, N. H., et al. (2014). "An Endotoxin Tolerance Signature Predicts Sepsis and Organ Dysfunction at Initial Clinical Presentation". *EBioMedicine* 1: p. 64.



- Pena, O. M., Pistolic, J., Raj, D., et al. (2011). "Endotoxin Tolerance Represents a Distinctive State of Alternative Polarization (M2) in Human Mononuclear Cells". *The Journal of Immunology* 186: pp. 7243–7254.
- Philip, A. M., Wang, Y., Mauro, A., et al. (2017). "Development of a Zebrafish Sepsis Model for High-Throughput Drug Discovery". *Molecular Medicine* 23: p. 134.
- Philippart, F., Fitting, C., and Cavaillon, J. M. (2012). "Lung microenvironment contributes to the resistance of alveolar macrophages to develop tolerance to endotoxin\*". *Critical care medicine* 40: pp. 2987–2996.
- Pichlmair, A., Schulz, O., Tan, C. P., et al. (2006). "RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates". *Science (New York, N.Y.)* 314: pp. 997–1001.
- Pillay, J., Den Braber, I., Vrisekoop, N., et al. (2010). "In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days". *Blood* 116: pp. 625–627.
- Pinder, e. M., Rostron, a. J., Hellyer, t. P., et al. (2018). "Randomised controlled trial of GM-CSF in critically ill patients with impaired neutrophil phagocytosis Critical care". *Thorax* 73: pp. 918–925.
- Platten, M., Nollen, E. A., Röhrig, U. F., et al. (2019). "Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond". *Nature reviews. Drug discovery* 18: pp. 379–401.
- Poll, T. van der, Shankar-Hari, M., and Wiersinga, W. J. (2021). "The immunology of sepsis". *Immunity* 54: pp. 2450–2464.
- Pollack, M. M., Patel, K. M., and Ruttimann, U. E. (1997). "The Pediatric Risk of Mortality III–Acute Physiology Score (PRISM III-APS): a method of assessing physiologic instability for pediatric intensive care unit patients". *The Journal of pediatrics* 131: pp. 575–581.
- Powell, T. J., Peng, Y., Berthoud, T. K., et al. (2013). "Examination of Influenza Specific T Cell Responses after Influenza Virus Challenge in Individuals Vaccinated with MVA-NP+M1 Vaccine". *PLOS ONE* 8: e62778.
- Prescott, H. C., Osterholzer, J. J., Langa, K. M., et al. (2016). "Late mortality after sepsis: propensity matched cohort study". *BMJ (Clinical research ed.)* 353.
- Qi, X., Yu, Y., Sun, R., et al. (2021). "Identification and characterization of neutrophil heterogeneity in sepsis". *Critical Care* 25.
- Quintelier, K., Couckuyt, A., Emmaneel, A., et al. (2021). "Analyzing high-dimensional cytometry data using FlowSOM". *Nature Protocols* 2021 16:8 16: pp. 3775–3801.
- Rahmel, T., Schmitz, S., Nowak, H., et al. (2020). "Long-term mortality and outcome in hospital survivors of septic shock, sepsis, and severe infections: The importance of aftercare". *PLOS ONE* 15: e0228952.
- Rajsbaum, R., Albrecht, R. A., Wang, M. K., et al. (2012). "Species-specific inhibition of RIG-I ubiquitination and IFN induction by the influenza A virus NS1 protein". *PLoS pathogens* 8.
- Rautanen, A., Mills, T. C., Gordon, A. C., et al. (2015). "Genome-wide association study of survival from sepsis due to pneumonia: An observational cohort study". *The Lancet Respiratory Medicine* 3: pp. 53–60.

- Reddy, K., Sinha, P., O’Kane, C. M., et al. (2020). “Subphenotypes in critical care: translation into clinical practice”. *The Lancet. Respiratory medicine* 8: pp. 631–643.
- Regev, A., Teichmann, S. A., Lander, E. S., et al. (2017). “The Human Cell Atlas”. *eLife* 6.
- Reis, E. S., Mastellos, D. C., Hajishengallis, G., et al. (2019). “New insights into the immune functions of complement”. *Nature Reviews Immunology* 2019 19:8 19: pp. 503–516.
- Ren, X., Wen, W., Fan, X., et al. (2021). “COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas”. *Cell* 184: p. 1895.
- Reyes, M., Filbin, M. R., Bhattacharyya, R. P., et al. (2020). “An immune cell signature of bacterial sepsis”. *Nature medicine* 26: p. 333.
- Reyes, M., Filbin, M. R., Bhattacharyya, R. P., et al. (2021). “Plasma from patients with bacterial sepsis or severe COVID-19 induces suppressive myeloid cell production from hematopoietic progenitors in vitro”. *Science Translational Medicine* 13: p. 9599.
- Rhee, C., Jones, T. M., Hamad, Y., et al. (2019). “Prevalence, Underlying Causes, and Preventability of Sepsis-Associated Mortality in US Acute Care Hospitals”. *JAMA Network Open* 2: e187571–e187571.
- Ricci, Z., Polito, A., Polito, A., et al. (2011). “The implications and management of septic acute kidney injury”. *Nature Reviews Nephrology* 2011 7:4 7: pp. 218–225.
- Ricciotti, E. and Fitzgerald, G. A. (2011). “Prostaglandins and Inflammation”. *Arteriosclerosis, thrombosis, and vascular biology* 31: p. 986.
- Ritchie, M. E., Phipson, B., Wu, D., et al. (2015). “limma powers differential expression analyses for RNA-sequencing and microarray studies”. *Nucleic Acids Research* 43: e47–e47.
- Roberts, G. H. L., Partha, R., Rhead, B., et al. (2022). “Expanded COVID-19 phenotype definitions reveal distinct patterns of genetic association and protective effects”. *Nature Genetics* 2022 54:4 54: pp. 374–381.
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. *Bioinformatics* 26: p. 139.
- Röltgen, K., Nielsen, S. C., Silva, O., et al. (2022). “Immune imprinting, breadth of variant recognition, and germinal center response in human SARS-CoV-2 infection and vaccination”. *Cell* 185: 1025–1040.e14.
- Root, R. K., Lodato, R. F., Patrick, W., et al. (2003). “Multicenter, double-blind, placebo-controlled study of the use of filgrastim in patients hospitalized with pneumonia and severe sepsis”. *Critical care medicine* 31: pp. 367–373.
- Rosas-Ballina, M., Ochani, M., Parrish, W. R., et al. (2008). “Splenic nerve is required for cholinergic antiinflammatory pathway control of TNF in endotoxemia”. *Proceedings of the National Academy of Sciences of the United States of America* 105: pp. 11008–11013.
- Rosier, F., Brisebarre, A., Dupuis, C., et al. (2021). “Genetic Predisposition to the Mortality in Septic Shock Patients: From GWAS to the Identification of a Regulatory Variant Modulating the Activity of a CISH Enhancer”. *International Journal of Molecular Sciences* 22.

- Rousseeuw, P. J. (1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". *Journal of Computational and Applied Mathematics* 20: pp. 53–65.
- Rudd, K. E., Johnson, S. C., Agesa, K. M., et al. (2020). "Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study". *The Lancet* 395: pp. 200–211.
- Rydzynski Moderbacher, C., Ramirez, S. I., Dan, J. M., et al. (2020). "Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity". *Cell* 183: 996–1012.e19.
- Salama, C., Han, J., Yau, L., et al. (2021). "Tocilizumab in Patients Hospitalized with Covid-19 Pneumonia". *New England Journal of Medicine* 384: pp. 20–30.
- Sander, L. E., Sackett, S. D., Dierssen, U., et al. (2010). "Hepatic acute-phase proteins control innate immune responses during infection by promoting myeloid-derived suppressor cell function". *The Journal of experimental medicine* 207: pp. 1453–1464.
- Sarkar, A. and Stephens, M. (2021). "Separating measurement and expression models clarifies confusion in single-cell RNA sequencing analysis". *Nature genetics* 53: p. 770.
- Sato, R. and Nasu, M. (2015). "A review of sepsis-induced cardiomyopathy". *Journal of Intensive Care* 3.
- Scherag, A., Schönebeck, F., Kesselmeier, M., et al. (2016). "Genetic Factors of the Disease Course after Sepsis: A Genome-Wide Study for 28Day Mortality". *EBioMedicine* 12: pp. 239–246.
- Schmid, M. A., Takizawa, H., Baumjohann, D. R., et al. (2011). "Bone marrow dendritic cell progenitors sense pathogens via Toll-like receptors and subsequently migrate to inflamed lymph nodes". *Blood* 118: pp. 4829–4840.
- Schönebeck, F., Kuhnt, E., Scholz, M., et al. (2015). "Common genomic variation in the FER gene: useful to stratify patients with sepsis due to pneumonia?" *Intensive care medicine* 41: pp. 1379–1381.
- Schork, N. J., Murray, S. S., Frazer, K. A., et al. (2009). "Common vs. rare allele hypotheses for complex diseases". *Current opinion in genetics development* 19: pp. 212–219.
- Schulte-Schrepping, J., Reusch, N., Paclik, D., et al. (2020). "Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment". *Cell* 182: p. 1419.
- Scicluna, B. P., Vught, L. A. van, Zwinderman, A. H., et al. (2017). "Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study". *The Lancet. Respiratory medicine* 5: pp. 816–826.
- Semerad, C. L., Christopher, M. J., Liu, F., et al. (2005). "G-CSF potently inhibits osteoblast activity and CXCL12 mRNA expression in the bone marrow". *Blood* 106: pp. 3020–3027.
- Semerad, C. L., Liu, F., Gregory, A. D., et al. (2002). "G-CSF is an essential regulator of neutrophil trafficking from the bone marrow to the blood". *Immunity* 17: pp. 413–423.
- Sen, D. R., Kaminski, J., Barnitz, R. A., et al. (2016). "The epigenetic landscape of T cell exhaustion". *Science* 354: pp. 1165–1169.



- Seshi, B., Kumar, S., and King, D. (2003). "Multilineage gene expression in human bone marrow stromal cells as evidenced by single-cell microarray analysis". *Blood Cells, Molecules, and Diseases* 31: pp. 268–285.
- Sette, A. and Crotty, S. (2021). "Adaptive immunity to SARS-CoV-2 and COVID-19". *Cell* 184: p. 861.
- Seymour, C. W., Gomez, H., Chang, C. C. H., et al. (2017). "Precision medicine for all? Challenges and opportunities for a precision medicine approach to critical illness". *Critical Care* 21: pp. 1–11.
- Shah, S. J., Barish, P. N., Prasad, P. A., et al. (2020). "Clinical features, diagnostics, and outcomes of patients presenting with acute respiratory illness: A retrospective cohort study of patients with and without COVID-19". *eClinicalMedicine* 27.
- Shaheen, S., Fawaz, F., Shah, S., et al. (2018). "Differential Expression and Pathway Analysis in Drug-Resistant Triple-Negative Breast Cancer Cell Lines Using RNASeq Analysis". *International Journal of Molecular Sciences* 19.
- Shakoory, B., Carcillo, J. A., Chatham, W. W., et al. (2016). "Interleukin-1 Receptor Blockade Is Associated With Reduced Mortality in Sepsis Patients With Features of Macrophage Activation Syndrome: Reanalysis of a Prior Phase III Trial". *Critical care medicine* 44: pp. 275–281.
- Shalova, I. N., Lim, J. Y., Chittezhath, M., et al. (2015). "Human monocytes undergo functional re-programming during sepsis mediated by hypoxia-inducible factor-1 $\alpha$ ". *Immunity* 42: pp. 484–498.
- Shankar-Hari, M., Datta, D., Wilson, J., et al. (2018). "Early PREdiction of sepsis using leukocyte surface biomarkers: the ExPRES-sepsis cohort study". *Intensive care medicine* 44: pp. 1836–1848.
- Shankar-Hari, M., Fear, D., Lavender, P., et al. (2017). "Activation-Associated Accelerated Apoptosis of Memory B Cells in Critically Ill Patients With Sepsis". *Critical care medicine* 45: pp. 875–882.
- Shankar-Hari, M. and Rubenfeld, G. D. (2019). "Population enrichment for critical care trials: Phenotypes and differential outcomes". *Current Opinion in Critical Care* 25: pp. 489–497.
- Shankar-Hari, M., Saha, R., Wilson, J., et al. (2020). "Rate and risk factors for rehospitalisation in sepsis survivors: systematic review and meta-analysis". *Intensive care medicine* 46: pp. 619–636.
- Shelton, J. F., Shastri, A. J., Fletez-Brant, K., et al. (2022). "The UGT2A1/UGT2A2 locus is associated with COVID-19-related loss of smell or taste". *Nature Genetics* 2022 54:2 54: pp. 121–124.
- Shelton, J. F., Shastri, A. J., Ye, C., et al. (2021). "Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity". *Nature Genetics* 2021 53:6 53: pp. 801–808.
- Silvin, A., Chapuis, N., Dunsmore, G., et al. (2020). "Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19". *Cell* 182: p. 1401.
- Sims, J. E. and Smith, D. E. (2010). "The IL-1 family: regulators of immunity". *Nature Reviews Immunology* 2010 10:2 10: pp. 89–102.
- Singer, M. (2008). "Cellular dysfunction in sepsis". *Clinics in chest medicine* 29: pp. 655–660.

- Singer, M. (2017). "Critical illness and flat batteries". *Critical Care* 21: pp. 69–73.
- Singer, M., Deutschman, C. S., Seymour, C., et al. (2016). "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)". *JAMA* 315: pp. 801–810.
- Singer, P. M., De Santis, V., Vitale, D., et al. (2004). "Multiorgan failure is an adaptive, endocrine-mediated, metabolic response to overwhelming systemic inflammation". *Lancet (London, England)* 364: pp. 545–548.
- Sinha, P., Matthay, M. A., and Calfee, C. S. (2020). "Is a "Cytokine Storm" Relevant to COVID-19?" *JAMA Internal Medicine* 180: pp. 1152–1154.
- Sinha, S., Rosin, N. L., Arora, R., et al. (2021). "Dexamethasone modulates immature neutrophils and interferon programming in severe COVID-19". *Nature Medicine* 2021 28:1 28: pp. 201–211.
- Smith, C. M., Wilson, N. S., Waithman, J., et al. (2004). "Cognate CD4+ T cell licensing of dendritic cells in CD8+ T cell immunity". *Nature Immunology* 2004 5:11 5: pp. 1143–1148.
- Smith, P. D., Suffredini, A. F., Allen, J. B., et al. (1994). "Endotoxin administration to humans primes alveolar macrophages for increased production of inflammatory mediators". *Journal of clinical immunology* 14: pp. 141–148.
- Sneller, M. C., Liang, C. J., Marques, A. R., et al. (2022). "A Longitudinal Study of COVID-19 Sequelae and Immunity: Baseline Findings". *Annals of Internal Medicine*.
- Sørensen, T. I., Nielsen, G. G., Andersen, P. K., et al. (1988). "Genetic and environmental influences on premature death in adult adoptees". *The New England journal of medicine* 318: pp. 727–732.
- Spitzer, M. H. and Nolan, G. P. (2016). "Mass Cytometry: Single Cells, Many Features". *Cell* 165: p. 780.
- Sposito, B., Broggi, A., Pandolfi, L., et al. (2021). "The interferon landscape along the respiratory tract impacts the severity of COVID-19". *Cell* 184: p. 4953.
- Stanski, N. L. and Wong, H. R. (2019). "Prognostic and predictive enrichment in sepsis". *Nature Reviews Nephrology* 2019 16:1 16: pp. 20–31.
- Stanzani, G., Tidswell, R., and Singer, M. (2020). "Do critical care patients hibernate? Theoretical support for less is more". *Intensive care medicine* 46: pp. 495–497.
- Stark, R., Grzelak, M., and Hadfield, J. (2019). "RNA sequencing: the teenage years". *Nature Reviews Genetics* 2019 20:11 20: pp. 631–656.
- Stein-O'Brien, G. L., Arora, R., Culhane, A. C., et al. (2018). "Enter the Matrix: Factorization Uncovers Knowledge from Omics". *Trends in Genetics* 34: pp. 790–805.
- Stephens, D. P., Thomas, J. H., Higgins, A., et al. (2008). "Randomized, double-blind, placebo-controlled trial of granulocyte colony-stimulating factor in patients with septic shock". *Critical care medicine* 36: pp. 448–454.
- Stephenson, E., Reynolds, G., Botting, R. A., et al. (2021). "Single-cell multi-omics analysis of the immune response in COVID-19". *Nature Medicine* 2021 27:5 27: pp. 904–916.
- Stewart, P. A., Luks, J., Roycik, M. D., et al. (2013). "Differentially Expressed Transcripts and Dysregulated Signaling Pathways and Networks in African American Breast Cancer". *PLoS ONE* 8: p. 82460.

- Stoeckius, M., Hafemeister, C., Stephenson, W., et al. (2017). "Simultaneous epitope and transcriptome measurement in single cells". *Nature Methods* 2017 14:9 14: pp. 865–868.
- Street, K., Risso, D., Fletcher, R. B., et al. (2018). "Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics". *BMC Genomics* 19: pp. 1–16.
- Su, S., Wong, G., Shi, W., et al. (2016). "Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses". *Trends in Microbiology* 24: p. 490.
- Subramanian, A., Tamayo, P., Mootha, V. K., et al. (2005). "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". *Proceedings of the National Academy of Sciences of the United States of America* 102: pp. 15545–15550.
- Summers, C., Rankin, S. M., Condliffe, A. M., et al. (2010). "Neutrophil kinetics in health and disease". *Trends in Immunology* 31: p. 318.
- Sun, J. C. and Bevan, M. J. (2003). "Defective CD8 T cell memory following acute infection without CD4 T cell help". *Science (New York, N.Y.)* 300: pp. 339–342.
- Sungnak, W., Huang, N., Bécavin, C., et al. (2020). "SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes". *Nature Medicine* 2020 26:5 26: pp. 681–687.
- Sutton, G. J., Poppe, D., Simmons, R. K., et al. (2022). "Comprehensive evaluation of deconvolution methods for human brain gene expression". *Nature Communications* 2022 13:1 13: pp. 1–18.
- Suzuki, T., Shimizu, T., Szalay, L., et al. (2006). "Androstenediol ameliorates alterations in immune cells cytokine production capacity in a two-hit model of trauma-hemorrhage and sepsis". *Cytokine* 34: pp. 76–84.
- Svensson, V. (2020). "Droplet scRNA-seq is not zero-inflated". *Nature Biotechnology* 2020 38:2 38: pp. 147–150.
- Sweeney, T. E., Azad, T. D., Donato, M., et al. (2018). "Unsupervised Analysis of Transcriptomics in Bacterial Sepsis Across Multiple Datasets Reveals Three Robust Clusters". *Critical care medicine* 46: pp. 915–925.
- Sweeney, T. E., Wong, H. R., and Khatri, P. (2016). "Robust classification of bacterial and viral infections via integrated host gene expression diagnostics". *Science translational medicine* 8.
- Tabula Sapiens Consortium\*, Jones, R. C., Karkanias, J., et al. (2022). "The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans." *Science (New York, N.Y.)* 376: eabl4896.
- Takahashi, T., Ellingson, M. K., Wong, P., et al. (2020). "Sex differences in immune responses that underlie COVID-19 disease outcomes". *Nature* 2020 588:7837 588: pp. 315–320.
- Takeuchi, O. and Akira, S. (2010). "Pattern Recognition Receptors and Inflammation". *Cell* 140: pp. 805–820.
- Tam, V., Patel, N., Turcotte, M., et al. (2019). "Benefits and limitations of genome-wide association studies". *Nature Reviews Genetics* 2019 20:8 20: pp. 467–484.
- Tang, F., Barbacioru, C., Wang, Y., et al. (2009). "mRNA-Seq whole-transcriptome analysis of a single cell". *Nature methods* 6: pp. 377–382.
- Tebartz, C., Horst, S. A., Sparwasser, T., et al. (2015). "A Major Role for Myeloid-Derived Suppressor Cells and a Minor Role for Regulatory T Cells

- in Immunosuppression during Staphylococcus aureus Infection". *The Journal of Immunology* 194: pp. 1100–1111.
- Thair, S. A., He, Y. D., Hasin-Brumshtein, Y., et al. (2021). "Transcriptomic similarities and differences in host response between SARS-CoV-2 and other viral infections". *iScience* 24: p. 101947.
- The RECOVERY Collaborative Group (2021). "Dexamethasone in Hospitalized Patients with Covid-19". *The New England journal of medicine* 384: pp. 693–704.
- The RECOVERY Collaborative Group (2022). "Baricitinib in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial and updated meta-analysis". *medRxiv*: p. 2022.03.02.22271623.
- Tietjen, I., Rihel, J. M., Cao, Y., et al. (2003). "Single-Cell Transcriptional Analysis of Neuronal Progenitors". *Neuron* 38: pp. 161–175.
- Tracey, K. J. (2002). "The inflammatory reflex". *Nature* 420: pp. 853–859.
- Tran, D., Nguyen, H., Tran, B., et al. (2021). "Fast and precise single-cell data analysis using a hierarchical autoencoder". *Nature Communications* 2021 12:1 12: pp. 1–10.
- Trede, N. S., Langenau, D. M., Traver, D., et al. (2004). "The Use of Zebrafish to Understand Immunity". *Immunity* 20: pp. 367–379.
- Trubetskoy, V., Pardiñas, A. F., Qi, T., et al. (2022). "Mapping genomic loci implicates genes and synaptic biology in schizophrenia". *Nature* 2022 604:7906 604: pp. 502–508.
- Trzuppek, D., Dunstan, M., Cutler, A. J., et al. (2020). "Discovery of CD80 and CD86 as recent activation markers on regulatory T cells by protein-RNA single-cell analysis". *Genome Medicine* 12.
- Uffelmann, E., Huang, Q. Q., Munung, N. S., et al. (2021). "Genome-wide association studies". *Nature Reviews Methods Primers* 2021 1:1 1: pp. 1–21.
- Uhel, F., Azzaoui, I., Grégoire, M., et al. (2017). "Early expansion of circulating granulocytic myeloid-derived suppressor cells predicts development of nosocomial infections in patients with sepsis". *American Journal of Respiratory and Critical Care Medicine* 196: pp. 315–327.
- Umansky, V., Blattner, C., Gebhardt, C., et al. (2016). "The Role of Myeloid-Derived Suppressor Cells (MDSC) in Cancer Progression". *Vaccines* 2016, Vol. 4, Page 36 4: p. 36.
- Vahedi, H. S. M., Bagheri, A., Jahanshir, A., et al. (2019). "Association of Lymphopenia with Short Term Outcomes of Sepsis Patients; a Brief Report". *Archives of Academic Emergency Medicine* 7: e14.
- Valenta, H., Erard, M., Dupré-Crochet, S., et al. (2020). "The NADPH Oxidase and the Phagosome". *Advances in experimental medicine and biology* 1246: pp. 153–177.
- Van Gelder, R. N., Von Zastrow, M. E., Yool, A., et al. (1990). "Amplified RNA synthesized from limited quantities of heterogeneous cDNA". *Proceedings of the National Academy of Sciences of the United States of America* 87: pp. 1663–1667.
- Van Vught, L. A., Klouwenberg, P. M., Spitoni, C., et al. (2016). "Incidence, Risk Factors, and Attributable Mortality of Secondary Infections in the Intensive Care Unit After Admission for Sepsis". *JAMA* 315: pp. 1469–1479.
- Veglia, F., Perego, M., and Gabrilovich, D. (2018). "Myeloid-derived suppressor cells coming of age". *Nature Immunology* 2018 19:2 19: pp. 108–119.

- Veglia, F., Sanseviero, E., and Gabrilovich, D. I. (2021). "Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity". *Nature Reviews Immunology* 2021 21:8 21: pp. 485–498.
- Verity, R., Okell, L. C., Dorigatti, I., et al. (2020). "Estimates of the severity of coronavirus disease 2019: a model-based analysis". *The Lancet Infectious Diseases* 20: pp. 669–677.
- Villani, A. C., Satija, R., Reynolds, G., et al. (2017). "Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors". *Science (New York, N.Y.)* 356.
- Wagar, L. E., Gentleman, B., Pircher, H., et al. (2011). "Influenza-Specific T Cells from Older People Are Enriched in the Late Effector Subset and Their Presence Inversely Correlates with Vaccine Response". *PLOS ONE* 6: e23698.
- Waldman, A. D., Fritz, J. M., and Lenardo, M. J. (2020). "A guide to cancer immunotherapy: from T cell basic science to clinical practice". *Nature Reviews Immunology* 2020 20:11 20: pp. 651–668.
- Walker, F., Zhang, H. H., Matthews, V., et al. (2008). "IL6/sIL6R complex contributes to emergency granulopoietic responses in G-CSF- and GM-CSF-deficient mice". *Blood* 111: pp. 3978–3985.
- Walton, A. H., Muenzer, J. T., Rasche, D., et al. (2014). "Reactivation of Multiple Viruses in Patients with Sepsis". *PLOS ONE* 9: e98819.
- Wang, T., Derhovanessian, A., De Cruz, S., et al. (2014). "Subsequent infections in survivors of sepsis: epidemiology and outcomes". *Journal of intensive care medicine* 29: pp. 87–95.
- Wang, X., Park, J., Susztak, K., et al. (2019). "Bulk tissue cell type deconvolution with multi-subject single-cell expression reference". *Nature Communications* 2019 10:1 10: pp. 1–9.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). "RNA-Seq: a revolutionary tool for transcriptomics". *Nature reviews. Genetics* 10: p. 57.
- Washburn, M. L., Wang, Z., Walton, A. H., et al. (2019). "T Cell- and Monocyte-Specific RNA-Sequencing Analysis in Septic and Nonseptic Critically Ill Patients and in Patients with Cancer". *Journal of immunology (Baltimore, Md. : 1950)* 203: pp. 1897–1908.
- Wei, Z., Li, C., Zhang, Y., et al. (2020). "Macrophage-Derived IL-1 $\beta$  Regulates Emergency Myelopoiesis via the NF- $\kappa$ B and C/ebp $\beta$  in Zebrafish". *Journal of immunology (Baltimore, Md. : 1950)* 205: pp. 2694–2706.
- Wen, M., Cai, G., Ye, J., et al. (2020). "Single-cell transcriptomics reveals the alteration of peripheral blood mononuclear cells driven by sepsis". *Annals of Translational Medicine* 8: pp. 125–125.
- Wiersinga, W. J., Leopold, S. J., Cranendonk, D. R., et al. (2013). "Host innate immune responses to sepsis". <https://doi.org/10.4161/viru.25436> 5: pp. 36–44.
- Wigerblad, G., Cao, Q., Brooks, S., et al. (2022). "Single-cell analysis reveals the range of transcriptional states of circulating human neutrophils". *bioRxiv*: p. 2022.02.22.481522.
- Wilk, A. J., Rustagi, A., Zhao, N. Q., et al. (2020). "A single-cell atlas of the peripheral immune response in patients with severe COVID-19". *Nature Medicine* 2020 26:7 26: pp. 1070–1076.



- Wilkerson, M. D. and Hayes, D. N. (2010). "ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking". *Bioinformatics* 26: p. 1572.
- Williamson, E. J., Walker, A. J., Bhaskaran, K., et al. (2020). "Factors associated with COVID-19-related death using OpenSAFELY". *Nature* 2020 584:7821 584: pp. 430–436.
- Wissink, E. M., Vihervaara, A., Tippens, N. D., et al. (2019). "Nascent RNA Analyses: Tracking Transcription and Its Regulation". *Nature reviews. Genetics* 20: p. 705.
- Witten, D. M. and Tibshirani, R. J. (2009). "Extensions of Sparse Canonical Correlation Analysis with Applications to Genomic Data". *Statistical Applications in Genetics and Molecular Biology* 8.
- Wolf, F. A., Angerer, P., and Theis, F. J. (2018). "SCANPY: Large-scale single-cell gene expression data analysis". *Genome Biology* 19: pp. 1–5.
- Wolf, F. A., Hamey, F. K., Plass, M., et al. (2019). "PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells". *Genome Biology* 20: pp. 1–9.
- Wong, H. R., Cvijanovich, N., Lin, R., et al. (2009). "Identification of pediatric septic shock subclasses based on genome-wide expression profiling". *BMC medicine* 7.
- Wong, H. R., Cvijanovich, N. Z., Allen, G. L., et al. (2011). "Validation of a gene expression-based subclassification strategy for pediatric septic shock". *Critical care medicine* 39: pp. 2511–2517.
- Wong, H. R., Cvijanovich, N. Z., Anas, N., et al. (2015). "Developing a clinically feasible personalized medicine approach to pediatric septic shock". *American journal of respiratory and critical care medicine* 191: pp. 309–315.
- Wong, H. R., Sweeney, T. E., Hart, K. W., et al. (2017a). "Pediatric sepsis endotypes among adults with sepsis". *Critical care medicine* 45: e1289.
- Wong, H. R., Sweeney, T. E., and Lindsell, C. J. (2017b). "Simplification of a Septic Shock Endotyping Strategy for Clinical Application". *American journal of respiratory and critical care medicine* 195: pp. 263–265.
- Wong, H. R., Wheeler, D. S., Tegtmeier, K., et al. (2010). "Toward a clinically feasible gene expression-based subclassification strategy for septic shock: proof of concept". *Critical care medicine* 38: pp. 1955–1961.
- Wu, R., Dong, W., Cui, X., et al. (2007). "Ghrelin down-regulates proinflammatory cytokines in sepsis through activation of the vagus nerve". *Annals of surgery* 245: pp. 480–486.
- Xu, C., Lopez, R., Mehlman, E., et al. (2021). "Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models". *Molecular Systems Biology* 17: e9620.
- Xu, D. and Tian, Y. (2015). "A Comprehensive Survey of Clustering Algorithms". *Annals of Data Science* 2015 2:2 2: pp. 165–193.
- Yao, C., Bora, S. A., Parimon, T., et al. (2021). "Cell-Type-Specific Immune Dysregulation in Severely Ill COVID-19 Patients". *Cell Reports* 34: p. 108590.
- Yeung, M. L., Teng, J. L. L., Jia, L., et al. (2021). "Soluble ACE2-mediated cell entry of SARS-CoV-2 via interaction with proteins related to the renin-angiotensin system". *Cell* 184: p. 2212.

- Youn, J.-I., Collazo, M., Shalova, I. N., et al. (2012). "Characterization of the nature of granulocytic myeloid-derived suppressor cells in tumor-bearing mice". *Journal of Leukocyte Biology* 91: pp. 167–181.
- Youngs, J., Provine, N. M., Lim, N., et al. (2021). "Identification of immune correlates of fatal outcomes in critically ill COVID-19 patients". *PLOS Pathogens* 17: e1009804.
- Yu, G., Wang, L. G., Han, Y., et al. (2012). "clusterProfiler: an R package for comparing biological themes among gene clusters". *Omics : a journal of integrative biology* 16: pp. 284–287.
- Yu, J., Du, W., Yan, F., et al. (2013). "Myeloid-derived suppressor cells suppress antitumor immune responses through IDO expression and correlate with lymph node metastasis in patients with breast cancer". *Journal of immunology (Baltimore, Md. : 1950)* 190: pp. 3783–3797.
- Zaitsev, K., Bambouskova, M., Swain, A., et al. (2019). "Complete deconvolution of cellular mixtures based on linearity of transcriptional signatures". *Nature communications* 10.
- Zanetta, L., Marcus, S. G., Vasile, J., et al. (2000). "EXPRESSION OF VON WILLEBRAND FACTOR, AN ENDOTHELIAL CELL MARKER, IS UP-REGULATED BY ANGIOGENESIS FACTORS: A POTENTIAL METHOD FOR OBJECTIVE ASSESSMENT OF TUMOR ANGIOGENESIS". *J. Cancer* 85: pp. 281–288.
- Zhang, B. and Horvath, S. (2005). "A general framework for weighted gene co-expression network analysis". *Statistical applications in genetics and molecular biology* 4.
- Zhang, H., Nguyen-Jackson, H., Panopoulos, A. D., et al. (2010). "STAT3 controls myeloid progenitor growth during emergency granulopoiesis". *Blood* 116: pp. 2462–2471.
- Zhang, Q., Liu, Z., Moncada-Velez, M., et al. (2020). "Inborn errors of type I IFN immunity in patients with life-threatening COVID-19". *Science (New York, N.Y.)* 370.
- Zheng, G. X., Terry, J. M., Belgrader, P., et al. (2017). "Massively parallel digital transcriptional profiling of single cells". *Nature Communications* 2017 8:1 8: pp. 1–12.
- Zhou, Z., Ren, L., Zhang, L., et al. (2020). "Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients". *Cell Host and Microbe* 27: 883–890.e2.
- Zhu, L., Yang, P., Zhao, Y., et al. (2020). "Single-Cell Sequencing of Peripheral Mononuclear Cells Reveals Distinct Immune Response Landscapes of COVID-19 and Influenza Patients". *Immunity* 53: pp. 685–696.
- Zollner, A., Koch, R., Jukic, A., et al. (2022). "Postacute COVID-19 is Characterized by Gut Viral Antigen Persistence in Inflammatory Bowel Diseases". *Gastroenterology* S0016-5085: pp. 00450–4.